THE PROCEEDINGS OF THE PHYSICAL SOCIETY

FROM JANUARY 1935 TO NOVEMBER 1935

VOLUME 47

Published by
THE PHYSICAL SOCIETY
1 Lowther Gardens, Exhibition Road
London, S.W. 7

Printed at
THE UNIVERSITY PRESS, CAMBRIDGE

12303

PRINTED IN GREAT BRITAIN

OFFICERS AND COUNCIL, 1935–1936

President:

The Right Hon. Lord Rayleigh, M.A., LL.D., Sc.D., F.Inst.P., F.R.S.

Vice-Presidents:

Who have filled the Office of President.

Sir OLIVER J. LODGE, D.Sc., LL.D., F.Inst.P., F.R.S.
Sir RICHARD GLAZEBROOK, K.C.B., D.Sc., F.Inst.P., F.R.S.
Sir J. J. THOMSON, O.M., Sc.D., F.Inst.P., F.R.S.
Prof. Sir CHARLES VERNON BOYS, LL.D., F.Inst.P., F.R.S.
Prof. C. H. LEES, D.Sc., F.Inst.P., F.R.S.
Prof. Sir W. H. BRAGG, K.B.E., M.A., F.Inst.P., F.R.S.
ALEXANDER RUSSELL, M.A., D.Sc., F.Inst.P., F.R.S.
Sir FRANK E. SMITH, K.C.B., C.B.E., D.Sc., F.Inst.P., Sec. R.S.
Prof. O. W. RICHARDSON, M.A., D.Sc., F.R.S.
W. H. ECCLES, D.Sc., F.Inst.P., F.R.S.
Prof. Sir A. S. EDDINGTON, M.A., D.Sc., F.R.S.
Prof. A. O. RANKINE, O.B.E., D.Sc., F.Inst.P., F.R.S.

Vice-Presidents:

T. SMITH, M.A., F.Inst.P., F.R.S.
Prof. W. WILSON, D.Sc., Ph.D., F.R.S.
D. OWEN, B.A., D.Sc., F.Inst.P.
J. H. AWBERY, B.Sc., F.Inst.P.

Hon. Secretaries:

EZER GRIFFITHS, D.Sc., F.Inst.P., F.R.S.
ALLAN FERGUSON, M.A., D.Sc., F.Inst.P.

Hon. Foreign Secretary:

Prof. O. W. RICHARDSON, M.A., D.Sc., F.R.S.

Hon. Treasurer:

R. W. PAUL, M.I.E.E., F.Inst.P.

Hon. Librarian:

J. H. BRINKWORTH, D.Sc., A.R.C.S., F.Inst.P.

Ordinary Members of Council:

LEWIS SIMONS, D.Sc., F.Inst.P.
R. S. WHIPPLE, M.I.E.E., F.Inst.P.
Prof. J. A. CROWTHER, M.A., Sc.D., F.Inst.P.
Prof. G. I. FINCH, M.B.E., F.Inst.P.
H. SHAW, D.Sc., A.R.C.S., F.Inst.P.
Prof. H. R. ROBINSON, D.Sc., Ph.D., F.Inst.P., F.R.S.
Miss M. O. SALTMARSH, Ph.D., F.Inst.P.
L. F. BATES, B.Sc., Ph.D., F.Inst.P.
L. HARTSHORN, D.Sc., A.R.C.S.
Prof. G. F. J. TEMPLE, D.Sc., Ph.D., F.Inst.P.
Prof. A. F. C. POLLARD, A.R.C.S., F.Inst.P.
H. H. EMSLEY, B.Sc.

Editor of the Proceedings:

Capt. C. W. HUME, M.C., B.Sc.

CONTENTS

Part 1. *January* 1, 1935.

PAGE

W. E. Bentham. Electronic theory and the magnetron oscillator 1

H. R. Nettleton and E. G. Balls. The absolute measurement of electrical resistance by a new rotating-coil method 54

John J. Manley. A new precision colorimeter 69

F. Llewellyn Jones. The energy of agitation of positive ions in argon . . 74

N. F. Astbury. The computation of the integrals required in mutual-inductance calculations 86

Beryl M. Dent. On observations of points connected by a linear relation . 92

R. O. Jenkins. Oxide films on liquid metals studied by means of electron-diffraction 109

D. H. Follett. The use of microphotometric methods in divided-beam spectrophotometry 125

E. B. Pearson. On the behaviour of suspended particles in air, and the velocity of sound at supersonic frequencies 136

A. B. Wood and F. D. Smith. The velocity of sound in sheet materials . . 149

N. R. Campbell and G. C. Marris. The measurement of loudness 153

A. T. Starr. The rectifying peak voltmeter as a standard instrument . . 184

Demonstration of the velocity of sound in sheet material 185

Demonstration of a simple apparatus for measuring the pull between two magnetized surfaces 186

Reviews of books 191

Part 2. *March* 1, 1935.

L. F. Bates and D. V. Reddi Pantulu. The magnetic properties of amorphous manganese 197

W. D. Wright and F. H. G. Pitt. The colour-vision characteristics of two trichromats 205

L. H. G. Dines. The potential acquired in the natural electric field by a vertical rod standing on the ground, insulated at the bottom and carrying a collector at the top 218

R. K. Asundi, R. Samuel and M. Zaki-uddin. The band systems of cadmium fluoride 235

A. K. Sen Gupta. Rotational analysis of the ultra-violet bands of phosphorus monoxide 247

L. F. Richardson. Time-marking a cathode-ray oscillograph by harmonics . 258

G. Millington. Ionization charts of the upper atmosphere, Part II . . . 263

W. A. Leyshon. Some experiments on electronic oscillations 277

M. C. Johnson. Models of the superposition and interpenetration of components in gas mixtures adsorbed upon thermionic, photoelectric, and catalytic surfaces: Part I, principles 287

F. F. P. Bisacre. Convergent polarized light and Hertz's problem for a uniaxial material 306

D. F. Martyn. The propagation of medium radio waves in the ionosphere . 323

PAGE

D. F. MARTYN, R. O. CHERRY and A. L. GREEN. Long-distance observations of radio waves of medium frequencies 340

BRUCE CHALMERS. An interference extensometer and some observations on the elasticity of lead 352

Reviews of books 371

Part 3. May 1, 1935.

PROFESSOR L. S. PALMER and DENIS TAYLOR. On a theory of the action of rectangular short-wave frame aerials 377

PROFESSOR L. S. PALMER and ROY WITTY. The current variations in a short-wave square frame aerial revolving in its own plane 388

J. G. HOLMES. Rapid mathematical methods for trichromatic colorimetry . . . 400

N. THOMPSON. Note on additional experiments on the effective rotation temperature of the negative glow in nitrogen 413

A. HARVEY and H. BELL. The band spectrum of beryllium monoxide . . . 415

S. E. WILLIAMS. The efficiency of excitation of the nitrogen first positive bands by electron impact 420

A. J. MADDOCK. Absolute intensities in the spectrum of quartz mercury arcs and their variation with temperature-changes of the surrounding air 424

P. C. MAHANTI. The band spectrum of vanadium oxide 433

E. D. EYLES and E. W. H. SELWYN. New method of measuring the time and efficiency of photographic shutters 446

HAROLD JEFFREYS. Time and amplitude relations in seismology 455

R. S. BURDON. Adsorption of gases on mercury 460

C. W. OATLEY. A negative-resistance device and its application to harmonic analysis 471

G. P. BARNARD. A new selenium-sulphur rectifier photoelectric cell 477

W. TAYLOR and H. W. LEE. The development of the photographic lens . . 502

Demonstration of apparatus for measuring the viscosity of liquids at high pressures 519

Reviews of books 521

Part 4. July 1, 1935.

W. RAILSTON and E. G. RICHARDSON. The effect of pressure on supersonic dispersion in gases 533

H. L. PENMAN. The effect of temperature on supersonic dispersion in gases . 543

W. N. BOND. The surface tension of a moving water sheet 549

JEANNE LIQUIER-MILWARD. Magnetic susceptibility of cerium chloride in aqueous solution and its variation with temperature 559

N. F. MOTT. A discussion of the transition metals on the basis of quantum mechanics 571

S. WHITEHEAD and W. G. RADLEY. Experiments relating to the distribution of alternating electric currents in the earth and the measurement of the resistivity of the earth 589

A. J. WOODALL. The direct measurement of the Peltier coefficient 615

D. P. R. PETRIE. Comparison of X-ray wave-lengths by the plane-grating vacuum spectrograph, and the structure of the K line of carbon 626

PAGE

N. W. McLachlan and A. L. Meyers. Spherical sound-waves of finite amplitude 644

W. J. Sulston. The temperature variation of the viscosity of aqueous solutions of strong electrolytes 657

Prof. J. T. MacGregor-Morris and D. E. H. Jones. A spectrographic examination of arcs between plain soot carbons and its connection with the candle-power per ampere of the positive crater 667

F. Simon, A. H. Cooke and H. Pearson. Liquefaction of hydrogen by the expansion method 678

J. H. Awbery and Ezer Griffiths. An investigation of the wet-and-dry-bulb hygrometer at low temperatures 684

G. Burniston Brown. On vortex motion in gaseous jets and the origin of their sensitivity to sound 703

Bruce Chalmers. The twinning of single crystals of tin 733

Arthur H. Compton. The twentieth Guthrie Lecture: An attempt to analyse cosmic rays 747

Reviews of books 774

Discussion on paper by W. Railston and E. G. Richardson 777

Part 5. *September* 1, 1935.

A. B. Wood. A correction to the theory of the Rayleigh disc as applied to the measurement of sound-intensity in water 779

A. B. Wood. An experimental determination of the frequencies of free circular plates 794

Norman Campbell. The statistical theory of errors 800

R. Cockburn. The variation of voltage-distribution and of electron transit-time with current in the planar diode 801

C. J. B. Clews. The electrical conductivity of some strong electrolytes in dilute solution and its variation over the temperature range 18° C. to 85° C. . . 818

H. P. Barasch. An improved counting circuit 824

C. W. Bunn. The lattice-dimensions of zinc oxide 835

Professor J. Hollingworth. The structure of the ionosphere 843

T. H. Pi and William Band. The longitudinal thermoelectric effect: (2) Nickel in longitudinal magnetic fields 852

M. K. Li and William Band. The longitudinal thermoelectric effect: (3) Aluminium 859

William Band. The longitudinal thermoelectric effect: (4) A further study of aluminium 862

H. J. Taylor. The disintegration of boron by neutrons 873

L. F. Bates and H. E. Hogwood. A note on the Raman spectrum of a ferromagnetic oxide 877

A. J. Bradley. The absorption factor for the powder and rotating-crystal methods of X-ray crystal analysis 879

G. F. C. Searle. The magnetic force at a point on its axis due to a current in a helical coil of one turn 900

J. L. Ch'en and William Band. Longitudinal thermoelectric effect: (5) silver . 904

Y. K. Hsü and William Band. Thermomagnetic hysteresis in nickel wire . 910

PAGE

R. L. SMITH-ROSE. The electrical properties of soil at frequencies up to 100 megacycles per second; with a note on the resistivity of ground in the United Kingdom 923

J. C. M. BRENTANO. The quantitative measurement of the intensity of X-ray reflections from crystalline powders 932

F. F. P. BISACRE. The theory of the formation of an image by a plane band grating used in the soft X-ray region 948

GEOFFREY D. PEGLER. A dynamometer null method of measuring the inductance and the effective resistance of iron-cored chokes carrying direct current . 964

S. WHITEHEAD and W. NETHERCOT. The breakdown of dielectrics under high voltage, with particular reference to thermal instability 974

Demonstration of the use of a liquid surface carrying ripples as a diffraction grating 998

Reviews of books 1000

Part 6. *November* 1, 1935.

PROFESSOR T. H. LABY and ASSOCIATE-PROFESSOR E. O. HERCUS. The effect of the aeration of the water used in the determination of the mechanical equivalent of heat 1003

J. S. PRESTON. The colour-matching of Tungsten-Filament lamps by means of a single photocell and colour filters 1012

R. H. SLOANE and C. M. MINNIS. Spectroscopic observations of recurrent phenomena in discharge tubes 1019

L. C. JACKSON. The magnetic moment of the manganic ion 1029

A. R. PEARSON and B. PLEASANCE. The colour temperatures of the Hefner and acetylene flames 1032

R. A. CHIPMAN. The electron-oscillation characteristics of an experimental plane-electrode triode 1042

T. H. ODDIE. The efficiency of separation of hydrogen and deuterium by electrolysis 1060

R. DONALDSON. A trichromatic colorimeter 1068

E. V. APPLETON and D. B. BOOHARIWALLA. The influence of a magnetic field on the high-frequency conductivity of an ionized medium 1074

GEOFFREY BUILDER and A. L. GREEN. Modulation-frequency-change technique for ionospheric measurements 1085

O. O. PULLEY. A receiver discriminating between right- and left-hand circularly polarized wireless waves 1098

H. GOULBOURNE JONES. Measurements of the thermal expansion of cast and rolled zinc 1117

Obituary Notices:

- F. J. WITTS 1129
- C. F. B. KEMP 1129
- SIR ARTHUR SCHUSTER 1130
- T. F. CONNOLLY 1134
- SIR ALFRED EWING 1135
- D. K. MORRIS 1136
- A. LYNCH 1137

Reviews of books 1139

Index to volume 47 1147

PROCEEDINGS AT THE MEETINGS OF THE PHYSICAL SOCIETY

SESSION 1934–35

Except where the contrary is stated the meetings were held at the Imperial College of Science and Technology, South Kensington, the President being in the Chair.

October 19, 1934.

The President announced that the Council had elected the following to Student Membership of the Society: Gordon Lennox James Bailey, G. W. T. Bird, Philip Sydney Brown, Ronald Bruce, Kenneth John Coppin, Mauekj Dhauji Dand, F. H. D. Eades, Edward Waddington Foster, William Gerald Harper, Kenneth Albert Harwood, Frank Holmes, Leonard House, Arnold Stanley Knight, Herbert Kolsky, George James Kynch, George Henry Lean, John Louis Michiels, Bryan Oliver Payne, Harry Plummer, Ivor William Ramsay, Thomas Henry Redding, M. A. Samad, Leslie Hamilton Tarrant, E. J. Whitmore.

The following presentation to the Joint Library was announced from the Chair as having been received since the last meeting; the thanks of the Society was accorded to the donor:

HOARE, F. E., *A Textbook of Thermodynamics*. Presented by the author.

The following papers were read:

"Velocity of sound in sheet materials. Chladni's figures at high frequencies," by A. B. WOOD, D.Sc., F.Inst.P. and F. D. SMITH.

"The measurement of loudness," by N. R. CAMPBELL, Sc.D., F.Inst.P. and G. C. MARRIS. *Followed by a discussion.*

A demonstration of simple apparatus for measuring the pull between two magnetized surfaces was given by Professor J. T. MACGREGOR-MORRIS, M.I.E.E. and Mr C. R. STONER, B.Sc. (Eng.), A.M.I.E.E.

November 2, 1934.

The following were elected to the Fellowship of the Society: Edmond van Aubel, Thomas Lydwell Eckersley, Leonard Frederick Ennever, William Francis Harling, Alfred Leonard Green, Maurice Milbourn, William Douglas Oliphant, Percy Watson Perryman, Basil Ferdinand Jamieson Schonland, Frank Arthur Short, Arthur Simons.

The following papers were read:

"A study of oxide films on liquid metals by electron-diffraction," by R. O. JENKINS, A.R.C.S., B.Sc.

"Electronic theory and the magnetron oscillator," by E. W. BENHAM.

The following papers were read in title:

"The absolute measurement of electrical resistance by a new rotating-coil method," by H. R. Nettleton, D.Sc. and E. G. Balls, M.C., B.Sc., A.I.C.

"A new precision tintometer," by J. J. Manley, M.A., D.Sc.

"The propagation of medium radio waves in the ionosphere," by D. F. Martyn, Ph.D., A.R.C.Sc.

"Long-distance observations of radio waves of medium frequencies," by D. F. Martyn, Ph.D., A.R.C.Sc., R. O. Cherry, M.Sc. and A. L. Green, Ph.D.

November 16, 1934.

The following were elected to the Fellowship of the Society: Charles Drummond Ellis, Donald Harry Smith, H. L. Brooke Woodyatt.

The President announced that Bhunthin Attagara and Keshab L. Chatterjee had been admitted to Student Membership of the Society.

The following papers were read:

"On the behaviour of suspended particles in air, and the velocity of sound at supersonic frequencies," by E. B. Pearson. (With demonstration.)

"The use of micro-photometric methods in divided beam spectrophotometry," by D. H. Follett, M.A., A.Inst.P.

The following papers were read in title:

"The energy of agitation of positive ions in argon," by F. Llewellyn Jones, M.A., D.Phil.

"The computation of the integrals required in mutual-inductance calculations," by N. F. Astbury, M.A.

"On observations of points connected by a linear graph," by Miss B. M. Dent, M.Sc.

A demonstration on the velocity of sound in sheet materials was given by A. B. Wood, D.Sc., F.Inst.P. and F. D. Smith.

December 7, 1934.

The following were elected to the Fellowship of the Society: Ursula Andrewes, Mark Benjamin, Robert Roscoe, William Stanley Mabe (transfer).

The following paper was read:

"An interference extensometer and some observations on the elasticity of lead," by Bruce Chalmers, B.Sc., Ph.D.

The following papers were read in title:

"The magnetic properties of amorphous manganese," by L. F. Bates, D.Sc., Ph.D., F.Inst.P. and D. V. Reddi Pantulu, B.Sc.

"Colour-vision characteristics of two trichromats," by W. D. Wright, D.Sc., Ph.D., A.R.C.S., D.I.C. and F. H. G. Pitt, A.R.C.S., D.I.C., B.Sc.

"On the potential acquired in the natural electric field by a vertical rod standing on the ground, insulated at the bottom and carrying a collector at the top," by L. H. G. Dines, M.A.

"The band systems of cadmium fluoride," by R. K. Asundi, R. Samuel and M. Zakiuddin.

"Rotational analysis of the ultra-violet bands of phosphorus monoxide," by A. K. Sen Gupta.

"Time-marking a cathode-ray oscillograph by means of harmonics," by L. F. Richardson, B.A., D.Sc., F.Inst.P., F.R.S.

The following demonstrations were given:

A demonstration of the properties of mumetal as applied in an earth inductor compass, by Captain O. E. Chapman.

A demonstration of apparatus for the measurement of the viscosity of liquids at high pressures, by C. C. Mason, O.B.E., M.A., F.Inst.P. (By the courtesy of the Burmah Oil Company, Ltd.)

December 21, 1934.

Meeting held at the Northampton Polytechnic Institute, St John Street, E.C. 1 by invitation of the Principal.

The following papers were read:

"Ionization charts of the upper atmosphere: Part II," by G. Millington, M.A., B.Sc.

"Some experiments on electronic oscillations," by W. A. Leyshon, Ph.D., F.Inst.P.

The following papers were read in title:

"Models of the superposition and interpenetration of components in gas mixtures adsorbed upon thermionic, photoelectric and catalytic surfaces: Part I, Principles," by M. C. Johnson, M.A., D.Sc., A.Inst.P.

"Adsorption of gases on mercury," by R. S. Burdon, B.Sc., F.Inst.P.

"Convergent polarized light and Hertz's problem for a uniaxial material," by F. F. P. Bisacre, M.A., B.Sc.

January 1, 2, 3, 1935.

The Twenty-fifth Annual Exhibition was held at the Imperial College of Science. The following discourses were delivered:

"The Architecture of Molecules," by B. Wheeler Robinson, M.A., Ph.D.

"The Problem of Ether Drift," by C. V. Drysdale, C.B., O.B.E., D.Sc., M.I.E.E., F.Inst.P.

"Giant Telescopes," by H. Spencer Jones, M.A., Sc.D., F.R.S.

January 18, 1935.

The following were elected to the Fellowship of the Society: Walter Betteridge, Leslie Giddens Brazier, Vivian Frederick Davey, W. E. Duncanson.

The President announced that the Council had elected James Henry Herbert Merriman to Student Membership of the Society.

Prizes and certificates awarded for the sixth competition in Craftsmanship and Draughtsmanship were presented.

An address was given by Mr W. Taylor, F.Inst.P., F.R.S., prepared by himself and Mr H. W. Lee, B.A., on "The Development of Photographic Lens Design."

February 1, 1935.

The following were elected to the Fellowship of the Society: P. M. S. Blackett, Charles Samuel Franklin, Henry Stafford Hatfield, A. W. Isenthal, Joseph Albert Lauwerys, Alexander D. McDonald, Charles Ernest Rickard, David A. Richards, George Reginald Stanbury, Ivor Willows Stray, Harold W. Thompson, F. W. G. White, Bernard George Whitmore, Dorothy Wrinch, Norman Lawrence Yates-Fish.

The Twentieth Guthrie Lecture was given by Professor A. H. Compton, of Chicago University, who took as his subject, "An attempt to analyse Cosmic Rays."

February 15, 1935.

The following were elected to the Fellowship of the Society: Richard Ablett, Herbert Langerman Crook, Kenneth Macfadyen, Harold Miller, David Evered Harrell Jones (transfer), Frederick Henry George Pitt (transfer).

The following papers were read:

"On a theory of the action of rectangular short-wave frame aerials," by L. S. Palmer, D.Sc., Ph.D., M.I.E.E., F.Inst.P. and Denis Taylor, M.Sc.

"On the current variations in a short-wave square frame aerial when revolving in its own plane," by L. S. Palmer, D.Sc., Ph.D., M.I.E.E., F.Inst.P. and R. Witty, B.Sc., A.Inst.P.

A demonstration of a relay that operates two circuits alternately with delayed action was given by D. S. Perfect, D.Sc.

March 1, 1935.

The following were elected to the Fellowship of the Society: Harold Albert Nancarrow, Eric Henry Dock (transfer).

The President announced that the Council had elected Gordon Edward Ashwell to Student Membership of the Society.

The following papers were read:

"Time and amplitude relations in seismology," by HAROLD JEFFREYS, F.R.S.

"A negative-resistance device and its application to harmonic analysis," by C. W. OATLEY, M.A., M.Sc.

The following papers were read in title:

"Rapid mathematical methods for trichromatic colorimetry," by J. G. HOLMES, A.R.C.S., B.Sc.

"Note on additional experiments on the effective rotation temperature of the negative glow in nitrogen," by N. THOMPSON, B.Sc.

"The band spectrum of beryllium monoxide," by A. HARVEY, Ph.D., F.Inst.P. and H. BELL, M.A.

"The efficiency of excitation of the nitrogen first positive bands by electron impact," by S. E. WILLIAMS, M.Sc.

The following demonstrations were given:

"Visible Frequency Valve-Instructor." A demonstration of some of the properties of a three-electrode valve at visible frequencies, by Professor H. F. TREWMAN, M.A., M.I.E.E.

"The C.R.O. Valve-Instructor." An application of the cathode ray oscillograph to instruction in the behaviour of thermionic valves in a variety of circuits, by G. F. NICHOLSON.

March 15, 1935.

Annual General Meeting.

The Minutes of the previous Annual General Meeting were read and accepted as correct.

The reports of the Council and Hon. Treasurer and the accounts were adopted.

The Officers and Council for 1935–6 and the Auditors were elected.

The Twelfth Duddell Medal was presented to Dr W. Ewart Williams.

Votes of thanks were accorded to the retiring Officers and Council and to the Governors of the Imperial College of Science and Technology.

Ordinary Meeting.

Robert Cockburn was elected to the Fellowship of the Society.

The President announced that the Council had elected Arthur Thomas Gill to Student Membership of the Society.

The following presentations to the Joint Library were announced from the Chair as having been received; the thanks of the Society were accorded to the donors:

Levy, H. and Baggott, E. A., *Numerical Studies in Differential Equations*. Vol. 1. Presented by Professor H. Levy.

Childs, W. H. J., *Physical Constants*. Presented by the author.

Philosophical Magazine, Nos. 214, 215. Presented by Major C. E. S. Phillips.

The following paper was read:

"The rectifier selenium-sulphur photoelectric cell," by G. P. Barnard, B.Sc., A.Inst.P.

The following papers were read in title:

"Absolute intensities in the spectrum of quartz mercury arcs and their variation with temperature changes of the surrounding air," by A. J. Maddock, M.Sc., A.Inst.P.

"The band spectrum of vanadium oxide," by P. C. Mahanti, M.Sc.

"A new method of measuring the time and efficiency of photographic shutters," by E. D. Eyles, B.Sc. and E. W. H. Selwyn, B.Sc., A.R.C.S., F.Inst.P.

April 5, 1935.

The following were elected to the Fellowship of the Society: Charles Alfred Coulson, Ifor Ceredig Jones, and Franz Simon.

The President announced that the Council had elected Madge G. Adam and Donald Martin Knight to Student Membership of the Society.

The following papers were read:

"The effect of pressure on supersonic dispersion in gases," by W. Railston, M.Sc. and E. G. Richardson, B.A., Ph.D., D.Sc.

"The effect of temperature on supersonic dispersion in gases," by H. L. Penman, M.Sc., A.Inst.P.

"The surface tension of a moving water sheet," by W. N. Bond, M.A., D.Sc., F.Inst.P. (With demonstration.)

The following paper was read in title:

"Magnetic susceptibility of cerium chloride in aqueous solution and its variation with temperature," by J. Liquier-Milward, D.Sc., Ph.D.

May 3, 1935.

John Stewart was elected to the Fellowship of the Society.

The following papers were read:

"Liquefaction of hydrogen by the expansion method," by F. Simon, A. H. Cooke and H. Pearson.

"On vortex motion in gaseous jets and the origin of their sensitivity to sound," by G. B. Brown, M.Sc., Ph.D. (With demonstration.)

"A discussion of the transition metals on the basis of quantum mechanics," by N. F. Mott, M.A.

The following papers were read in title:

"Experiments relating to the distribution of alternating electric currents in the earth and the measurement of earth resistivity," by S. Whitehead, M.A., Ph.D., A.M.I.E.E., F.Inst.P. and W. G. Radley, B.Sc.

"The quantitative measurement of the intensity of X-ray reflections from crystalline powders," by J. C. M. Brentano, D.Sc., F.Inst.P.

"The direct measurement of the Peltier coefficient," by A. J. Woodall, B.Sc., Ph.D., A.Inst.P.

"The longitudinal thermoelectric effect: (2) Nickel in longitudinal magnetic fields," by T. H. Pi, M.Sc. and W. Band, M.Sc.

"The longitudinal thermoelectric effect: (3) Aluminium," by M. K. Li, M.Sc. and W. Band, M.Sc.

"The longitudinal thermoelectric effect: (4) Aluminium (*continued*)," by W. Band, M.Sc.

May 17, 1935.

The following were elected to the Fellowship of the Society: David Forbes Martyn, Eugene Bloch.

The following presentations to the Joint Library were announced from the Chair as having been received; the thanks of the Society were accorded to the donors:

Journal of the Institution of Electrical Engineers, Vols. 64 and 65. Presented by Mr W. H. Nottage.

Journal of the Institution of Electrical Engineers, Vol. 66. Presented by Mr Albert Campbell.

The following papers were read:

"The twinning of single crystals of tin," by Bruce Chalmers, B.Sc., Ph.D.

"An investigation of the wet and dry bulb hygrometer at low temperatures," by J. H. Awbery, B.A., B.Sc., F.Inst.P. and Ezer Griffiths, D.Sc., F.Inst.P., F.R.S.

The following papers were read in title:

"Comparison of X-ray wave-lengths by the plane grating vacuum spectrograph and the structure of the K line of carbon," by D. P. R. Petrie, M.Sc.

"A spectrographic examination of arcs between plain soot carbons and its connection with the candle-power per ampere of the positive crater," by J. T. MacGregor-Morris, M.I.E.E. and D. E. H. Jones, M.Sc., A.Inst.P.

"Spherical sound waves of finite amplitude," by N. W. McLachlan, D.Sc., M.I.E.E. and A. L. Meyers, B.Sc., A.M.I.E.E.

"The temperature variation of the viscosity of aqueous solutions of strong electrolytes," by W. J. Sulston, B.Sc., A.Inst.P.

June 7, 1935.

Dorothy Jessie Alexander was elected to the Fellowship of the Society.

The President announced that the Council had elected the following to Student Membership of the Society: Arthur Hafford Cooke, Raymond Frederick Cyster, Samuel Silver, Abraham Yanovsky.

The following papers were read:

"A correction to the theory of the Rayleigh disc. Measurement of sound intensity in water," by A. B. Wood, D.Sc., F.Inst.P.

"Experimental determination of the frequencies of free circular plates," by A. B. Wood, D.Sc., F.Inst.P.

"The statistical theory of errors," by N. R. Campbell, Sc.D., F.Inst.P.

The following papers were read in title:

"The breakdown of dielectrics under high voltage, with particular reference to thermal instability," by S. Whitehead, M.A., Ph.D., A.M.I.E.E., F.Inst.P. and W. Nethercot, M.A., B.Sc.

"The variation of voltage distribution and of electron transit time with current in the planar diode," by R. Cockburn, B.Sc.

"The electrical conductivity of some strong electrolytes in dilute solution, and its variation over the temperature range 18° C. to 25° C.," by C. J. B. Clews, B.Sc., A.Inst.P.

"An improved counting circuit," by H. P. Barasch.

June 21, 1935.

The following papers were read:

"A null dynamometer method of measuring the inductance and the effective resistance of iron-cored chokes carrying direct current," by G. D. Pegler, B.Sc.

"The electrical properties of soil at frequencies up to 100 megacycles per second (with a note on the resistivity of ground in the United Kingdom)," by R. L. Smith-Rose, D.Sc., Ph.D., A.M.I.E.E.

The following papers were read in title:

"The lattice dimensions of zinc oxide," by C. W. Bunn.

"The structure of the ionosphere," by J. Hollingworth, M.A., D.Sc., M.I.E.E.

"The effect of the aeration of the water used in the determination of the mechanical equivalent of heat," by T. H. Laby, M.A., Sc.D., F.Inst.P., F.R.S. and E. O. Hercus, D.Sc., F.Inst.P.

"Longitudinal thermoelectric effect (5) Silver," by J. L. Ch'en, M.S. and W. Band, M.Sc.

"Thermomagnetic hysteresis of nickel," by Y. K. Hsü, M.S. and W. Band, M.Sc.

REPORT OF COUNCIL FOR THE YEAR ENDING FEBRUARY 28, 1935

MEETINGS

During the period under review 14 Ordinary Science Meetings were held at the Imperial College of Science and Technology. At these meetings 56 papers were presented and 7 demonstrations given.

A Science Meeting was held on December 21, 1934, at the Northampton Polytechnic Institute, by the kind invitation of the Principal, Mr S. C. Laws, and a number of demonstrations were given in the laboratories of the Institute.

On January 18, 1935, Mr William Taylor delivered a lecture prepared by himself and Mr H. W. Lee on "The Development of Photographic Lens Design".

SUMMER MEETING AT THE ROYAL NAVAL COLLEGE, GREENWICH

A meeting of the Society was held at the Royal Naval College, Greenwich, on June 16, 1934, by the kind invitation of the President, Vice-Admiral Sir B. E. Domville, and Professor G. B. Bryan.

The Painted Hall and other historical features were open to inspection by members and their guests.

A lecture on the history of the College was given by Professor G. A. R. Callendar in the forenoon and in the afternoon Professor Bryan delivered an interesting address dealing with certain anomalies and difficulties observed in lecture experiments.

In the Mechanics and Physical Laboratories experiments were shown by Professors Haigh and Bryan and their staffs. Dr L. J. Comrie gave a lecture, with demonstration, on some modern calculating, adding and printing machines.

GUTHRIE LECTURES

The Nineteenth Guthrie Lecture was delivered on May 4, 1934, by Sir Charles Vernon Boys on "My recent progress in Gas Calorimetry", and the Twentieth Guthrie Lecture was delivered on February 1, 1935, by Professor A. H. Compton of Chicago University on "An attempt to analyse Cosmic Rays".

INTERNATIONAL CONFERENCE

An International Conference on Physics, organized by the Royal Society and the Physical Society, was held in the week October 1–6, 1934, in conjunction with the meeting of the Union of Pure and Applied Physics. Meetings for the discussions of papers on Nuclear Physics and on the Solid State of Matter were held in London and Cambridge, those in London taking place either at the rooms of the Royal Society or in the Royal Institution. About 30 papers were contributed by authors from 10 different countries, and no less than 600 persons were enrolled as members of the Conference; 200 of these were visitors from foreign countries.

The proceedings of the Conference are being published by the Society, and Fellows will in due course receive particulars and order-forms.

THE DUDDELL MEDAL

The Eleventh Duddell Medal was presented to Mr H. Dennis Taylor on March 16, 1934, for his work on optical design.

The Council has awarded the Twelfth Duddell Medal to Dr W. Ewart Williams for his work on interferometry.

ANNUAL EXHIBITION

The Twenty-fifth Annual Exhibition was held on January 1, 2 and 3, 1935, at the Imperial College of Science and Technology by the courtesy of the Governing Body.

There were 79 exhibitors in the Trade Section and 32 in the Research and Experimental Section. The attendance during the three days was close on 9000.

The following discourses were given:

"The Architecture of Molecules", by B. Wheeler Robinson, M.A., Ph.D.

"The Problem of Ether Drift", by C. V. Drysdale, C.B., O.B.E., D.Sc., M.I.E.E., F.Inst.P.

"Giant Telescopes", by H. Spencer Jones, M.A., Sc.D., F.R.S.

PROGRESS REPORTS

The Society has for some time past had under consideration the possibility of issuing a series of annual progress reports on physics. The first volume of the series has just been issued and is now on sale. It contains reports on the various main branches of physics, together with a number of reports on special subjects. The volume runs to some 370 pages, the format being that of the *Proceedings*, and the price, bound in cloth, is 12*s*. 6*d*. to non-Fellows and 9*s*. 0*d*. to Fellows of the Society.

REPRESENTATION ON OTHER BODIES

The three members now representing the Physical Society on the British National Committee for Physics are Dr E. Griffiths, Mr T. Smith and Mr J. H. Awbery.

Professor A. Ferguson, Dr D. Owen, Professor G. F. J. Temple and Mr J. H. Awbery represent the Society on the Committee of Management of Science Abstracts.

Professor E. V. Appleton and Dr A. B. Wood are the representatives of the Society on the National Committee for Radio-Telegraphy.

Dr D. Owen and Dr E. Griffiths represent the Society on the Board of the Institute of Physics.

Dr W. M. Hampton is the Society's representative on the Council of the Fourth Glass Convention.

OBITUARY

The Council records with deep regret the deaths of the following Fellows: Madame Curie (Honorary Fellow), Professor A. P. Chattock, Mr T. F. Connolly, Sir Alfred Ewing, Professor W. M. Hicks, Mr C. F. B. Kemp, Mr A. Lynch, Rev. S. A. McDowall, Mr M. J. Salter, Sir Arthur Schuster, Mr G. P. Simpson, Professor the Rev. S. Sircom, Mr J. B. Styring, Dr F. Wenner, Mr F. J. Witts.

MEMBERSHIP ROLL AT DEC. 31, 1934

	Total Dec. 31, 1933	Changes during 1934				Total Dec. 31, 1934
Honorary Fellows	12	Deceased			1	11
Honorary Fellows (Optical Society)	8				–	8
Ex-officio Fellows	4				–	4
Ordinary Fellows	921	Elected		29		
		Student transfers ...		7		
				36		
		Deceased	11			
		Resigned or lapsed ...	25			
				36		
		Net change			–	921
Students	60	Elected		16		
		Transferred to Fellow	7			
		Resigned or lapsed ...	2			
				9		
		Net increase			7	67
Total Membership	1005	Net increase			6	1011

REPORT OF THE HONORARY TREASURER

THE accounts for the year ended December 31st, 1934 show an excess of income over expenditure of £95. 7*s*. 10*d*. The estimated value of the Society's publications has been shown in a footnote on the Balance Sheet.

The special account for the suggested publication of annual reports on the progress of physical science now stands at £142. 10*s*. 0*d*. As mentioned in the Report of the Council the first volume of the series has just been published.

The Society's investments have been valued at market prices on December 31st, 1934 through the courtesy of the Manager of the Charing Cross Branch of the Westminster Bank.

(Signed) ROBERT S. WHIPPLE
Honorary Treasurer

March 1*st*, 1935

INCOME AND EXPENDITURE ACCOUNT FOR THE YEAR ENDED 31ST DECEMBER, 1934

1933 £ s. d.	EXPENDITURE	£ s. d.	£ s. d.
	To		
535 0 0	"Science Abstracts"		534 0 0
	Normal Publications:		
1759 0 10	"Proceedings" and Exhibition Catalogue	1877 9 1	
106 11 3	Notices	56 9 0	
211 7 5	Editorial and General Expenses	222 12 8	2156 10 9
	Special Publications:		
	Printing and Dispatching "Report on Geometrical Optics"		106 12 10
283 16 3	Postage on Normal Publications and General Correspondence		272 16 9
61 2 1	Expenses at Meetings		34 4 4
13 5 0	Periodicals and Library		10 12 6
	Institute of Physics:		
500 0 0	Clerical and other Assistance	500 0 0	
296 2 3	Electricity, Labour and Sundries in connexion with Exhibition	297 11 10	797 11 10
112 10 4	Sundry Expenses		101 7 7
40 0 0	*Honoraria to Lecturers*		
24 3 0	*New Die for Duddell Medal*		
46 16 2	*Printing and Dispatching List of Members*		
512 1 9	Balance, being excess of Income over Expenditure, carried forward to Accumulated Fund		95 7 10
£4501 16 4			£4109 4 5

1933 £ s. d.	INCOME	£ s. d.	£ s. d.
	By		
	Subscriptions:		
35 14 0	Entrance Fees	27 6 0	
1645 2 4	Fellows*	1659 5 7	
4 4 0	Voluntary†	4 4 0	
28 17 0	Students	32 5 6	
108 6 1	For "Science Abstracts" and Advance Proofs	106 14 6	1829 15 7
96 0 0	Transfer from Life Composition Fund of amounts paid by Fellows now deceased		71 10 0
1072 10 10	Sales of Publications		825 14 6
324 13 9	Advertisements in Publications		374 14 6
	Dividends from Investments	261 19 0	
	Less Transfer to Duddell Memorial Fund	14 0 0	
236 5 1			247 19 0
850 3 3	Exhibitors' and other Payments on account of Exhibition		759 10 10
100 0 0	*Royal Society grant for Publications*		
£4501 16 4			£4109 4 5

* Seventy-nine Fellows paid reduced subscriptions by the arrangement with the Institute of Physics, the total rebate being £26. 16*s*. 2*d*.

† Voluntary subscriptions are subscriptions paid by Fellows who compounded for the low sum of £10.

BALANCE SHEET AS ON 31ST DECEMBER, 1934

BALANCE SHEET AS ON 31st DECEMBER, 1934

Liabilities

	£	s.	d.	£	s.	d.	£	s.	d.
Sundry Creditors				1420	17	8			
Life Compositions:									
As on December 31, 1933 . .	1622	0	0						
Add Payment during year . .	21	0	0						
	1643	0	0						
Less Transfer to Income and Expenditure Account of amounts paid by Fellows now deceased .	71	10	0						
				1571	10	0			
Subscriptions paid in advance:									
Members	63	9	11						
Publications	186	16	8						
				250	6	7			
							3242	14	3
TRUST FUNDS									
W. F. Stanley Trust Fund . .				334	0	0			
Duddell Memorial Trust Fund . .				459	12	9			
Optical Convention 1926, *Trust Account*				37	0	9			
Special Reports on Physics Account .				142	10	0			
							973	3	6
GENERAL FUND									
Accumulated Fund:									
As on December 31, 1933 . .	3180	1	0						
Add Increase in value of Investments	796	0	0						
Add Balance of Income and Expenditure Account . . .	95	7	10						
				4071	8	10			
A. W. Scott Bequest . . .				199	19	6			
							4271	8	4
							£8487	6	1

Assets

	£	s.	d.	£	s.	d.
Investments at Market Value on December 31, 1934:						
£1500 Consolidated Stock 2½%	1395	0	0			
£1000 War Loan 3½% Inscribed "A" Account .	1090	0	0			
£750 War Loan 3½% Registered	818	0	0			
£650 Funding Loan 4%	786	0	0			
£500 India 3½% Stock	495	0	0			
£400 Lancaster Corporation 3%	392	0	0			
£399 London Midland and Scottish Railway 4% Debenture Stock	431	0	0			
£1000 London Midland and Scottish Railway 4% Preference Stock	860	0	0			
£500 London and North Eastern Railway 4% Debenture Stock	540	0	0			
£150 Southern Railway 5% Debenture Stock .	198	0	0			
£300 Southern Railway Preferred Ordinary Stock .	237	0	0			
£442 Southern Railway Deferred Ordinary Stock .	97	0	0			
£400 War Loan 3½% Inscribed "B" Account .	436	0	0			
				7775	0	0
Dividends due from Investments				35	4	9
Inland Revenue—Income Tax claimed for 1934 . .				15	17	2
Subscriptions due				35	9	0
Sundry Debtors				115	7	0
Cash at Bank: Deposit	200	0	0			
Current	308	18	2			
Cash in Hand	1	10	0			
				510	8	2
				£8487	6	1
Stock of publications on 31*st December*, 1934. *At Honorary Treasurer's valuation*	£1395	3	3			

ROBERT S. WHIPPLE, *Honorary Treasurer.*

We have audited the above Balance Sheet and have obtained all the information and explanations we have required. We have verified the Bank Balances and the Investments. In our opinion such Balance Sheet is properly drawn up so as to exhibit a true and correct view of the state of the Society's affairs according to the best of our information and the explanations given to us and as shown by the books of the Society.

SPENCER HOUSE, SOUTH PLACE, E.C. 2
1st *March*, 1935.

KNOX, CROPPER & CO.,
Chartered Accountants.

LIFE COMPOSITION FUND ON DECEMBER 31ST, 1934

	£	s.	d.
55 Fellows paid £10	550	0	0
1 Fellow paid £15	15	0	0
15 Fellows paid £21	315	0	0
1 Fellow paid £30	30	0	0
21 Fellows paid £31. 10*s*.	661	10	0
	£1571	10	0

W. F. STANLEY TRUST FUND

	£	s.	d.		£	s.	d.
Carried to Balance Sheet	334	0	0	£300 Southern Railway Preferred Ordinary Stock	237	0	
				£442 Southern Railway Deferred Ordinary Stock	97	0	
	£334	0	0		£334	0	

DUDDELL MEMORIAL TRUST FUND

CAPITAL

	£	s.	d.		£	s.	d.
Carried to Balance Sheet	436	0	0	£400 War Loan 3½% Inscribed "B" Account	436	0	

REVENUE

	£	s.	d.		£	s.	d.
Honorarium to Medallist	15	0	0	Balance on December 31st, 1933	24	12	
Balance carried to Balance Sheet	23	12	9	Interest	14	0	
	£38	12	9		£38	12	

OPTICAL CONVENTION, 1926, TRUST ACCOUNT

	£	s.	d.		£	s.	d.
Balance carried to Balance Sheet	37	0	9	Balance on December 31st, 1933	31	18	
				Sales of Publications	5	2	
	£37	0	9		£37	0	

A. W. SCOTT BEQUEST

	£	s.	d.		£	s.	d.
Expenses in connexion with Library	12	5	7	Balance on December 31st, 1933	212	5	
Balance carried to Balance Sheet	199	19	6				
	£212	5	1		£212	5	

"SPECIAL REPORTS ON PHYSICS" ACCOUNT

	£	s.	d.		£	s.	d.
Balance carried to Balance Sheet	142	10	0	Balance on December 31st, 1933	142	2	
				Donations during 1934		8	
	£142	10	0		£142	10	

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 47, PART 1 January 1, 1935 No. 258

621.396.615.1 621.385.16

ELECTRONIC THEORY AND THE MAGNETRON OSCILLATOR

By W. E. BENHAM

Received August 9, 1934. Read November 2, 1934.

ABSTRACT. The present paper extends previous analysis* to cover any degree of space-charge limitation. In addition to terms involving initial velocities and accelerations of electrons, the effect of a magnetic field of constant value has been included.

The potential-distribution is shown to be but slightly influenced by the magnetic field under steady-state conditions. It is shown that Langmuir's results for this case† still apply fairly accurately for potential-minimum calculations in the presence of a magnetic field. Particular attention is given to the value of the electric field at the cathode and to the average initial velocity of those electrons which reach the potential-minimum and thus constitute the anode current. The prolongation of electron transit time by the magnetic field is fully discussed.

The electron-velocity is composed of a forward component parallel to the electric field and a drift component at right angles to both electric and magnetic fields. The forward velocity may be oscillatory in character but the drift velocity is unaffected by alternating fields.

The general solution for the forward velocity contains two arbitrary functions the correct choice of which is fully discussed. The inclusion of a magnetic field is definitely of assistance in establishing that the choice of the arbitrary functions is governed mainly by the necessity for avoiding infinite values, rather than by boundary conditions of the usual type. This anomalous state of affairs arises from the non-linearity of the original equations, which involves the introduction at an early stage of the arbitrarily assigned boundary values of the steady-state components of velocity and acceleration as well as those of the direct and alternating current.

Formulae are given for the components normal and parallel to the plates of the electron-velocity averaged over all electrons, and application is made to a real and to a virtual cathode.

§ 1. INTRODUCTION

In the vacuum-tube theory of some twenty years ago it was considered sufficient to know the relations existing between direct currents and voltages, as obtained theoretically or by direct measurement, to be well acquainted with the processes that take place in high-vacuum tubes. At the present day the same outlook is still very general, and not without abundant experimental justification. In the design of vacuum tubes for broadcast or commercial transmission and reception a number of constants have now, of course, to be determined in addition to the static characteristic curves. The most important of these are the electrostatic capacities between the several pairs of electrodes associated with the tube. Where curves other than the static characteristic curves are quoted, such as those showing audio output power and efficiency as a function of output impedance, such curves can generally be regarded as known from the static characteristics, though in practice dynamic measurements are usual where the saving of time is a major consideration.

---

* W. E. Benham, *Phil. Mag.* **11**, 457 (1931).

† *Phys. Rev.* **21**, 419 (1923).

It is generally true that the higher the frequency of the alternating current for which a given vacuum tube is designed the greater is the uncertainty that attaches to any estimate of the performance based on static characteristics. In addition to the limitations of efficiency associated with interelectrode capacities already referred to, dielectric and eddy-current losses may be paramount. In many cases the electron-inertia is rightly considered negligible in comparison with other factors. What is not generally appreciated, however, is that electron-inertia may be of importance even at audio frequencies. For example, the operating interelectrode capacities are known to depend on the space-charge conditions between the electrodes. If the electrons were possessed of charge but were fixed in space, they would invariably increase the capacity as compared with the electrostatic value. Owing to the finite mass of the electron the capacity is lessened by an amount depending on the space-current conditions and geometry of the electrode-pair considered, and only to a minor extent on the frequency (except at frequencies
T comparable with T^{-1}, where T is transit time of electrons).

It is, admittedly, difficult to predict with certainty the effect of space-charge inertia in multi-electrode valves. It is possible, however, to say with confidence that a stage is rapidly approaching where greater attention will have to be paid to this question. In the pentagrid tube, for example, we have a virtual cathode formed between a pair of grids and serving as electron-source to succeeding grids.

It appears that a virtual cathode is possessed of remarkable properties, accurate information concerning which can only be obtained by means of an exhaustive theoretical study covering electron-inertia, together with the initial velocities and accelerations at the real cathode.

In the attempt to establish some of the properties of a magnetron oscillator, a magnetic field applied parallel to the plates of a parallel plane diode was studied, first of all by neglecting initial velocities and accelerations, and anomalous results were obtained at the critical plane, which is the turning-point of electrons in the diode. Now the critical plane in a magnetron differs from a virtual cathode* only in respect of the existence of a drift-velocity component parallel to the plates. A study of the critical plane should therefore result in information which can be applied to virtual cathodes in general. With a view to obtaining further information, initial velocities and accelerations have been included, and the general solution has been derived.

---

* L. Tonks, *Phys. Rev.* **29**, 913 (1927).

§ 2. NATURE OF THE PROBLEM

(2.1) *General conceptions.* The problem involved in generalized electronic theory is very similar in nature to that of the motion of charged particles in free space under the influence of electromagnetic waves, but with one important difference. In illustration of this difference we may consider two equal electromagnetic waves travelling in opposite directions. Let them meet at a point. If the electric vectors of the two waves are parallel, a quasistationary condition, in which the magnetic forces in the waves cancel one another, will obtain at the point considered and also along the line of overlap of the waves. Further, if the waves considered have travelled sufficiently from their respective source to be considered plane, they will meet in a plane and a volume will be traced out in which the above quasistationary state applies.

If in this volume we insert a plane thermionic cathode in a condition to emit electrons, and orient the cathode at right angles to the electric vectors of the waves, the motions of the electrons in the space will be governed by the same considerations as apply in the practical case of a valve subjected to alternating potentials.

The illustration serves to show that the problem is somewhat simplified as compared with the general electromagnetic problem, which has hitherto defied solution, in that terms representing the magnetic vectors may be dropped out of the equations. This simplification is fortunate in that the complication is already considerable in view of the mutual repulsions of the electrons. Secondary magnetic effects due to the motions of the electrons themselves are small for velocities low compared with c, and these will be neglected along with the magnetic vectors, the only magnetic field considered being one of constant value.

Still another conception which it is desirable to introduce at an early stage is the subdivision into alternating and direct-current components of certain of the physical quantities appearing in the equations. This operation is effected in the ordinary course of the analysis, but a word as to its justification may not be out of place here. In previous work the vector quantities all occurred in the same direction. Any vector could therefore be subdivided into components of a Fourier series which could be added together at any point where desired. In the present case we have to justify the operation of choosing components of the physical quantities parallel to the axes of co-ordinates and afterwards effecting the harmonic subdivision. In this connection we have to deal with the non-linear expression $U\nabla . U$, which, when components are taken parallel to the axes, represents nine terms in all. While the velocity U of the electrons is itself a vector, $U\nabla$ denotes the scalar operator $U\nabla$

$$\left(U_x\frac{\partial}{\partial x}+U_y\frac{\partial}{\partial y}+U_z\frac{\partial}{\partial z}\right).$$

The x-component of $U\nabla . U$ is thus

$$\left(U_x\frac{\partial}{\partial x}+U_y\frac{\partial}{\partial y}+U_z\frac{\partial}{\partial z}\right)U_x.$$

If we effect harmonic subdivision before dividing into components parallel to the axes, we obtain ($\bar{U}$ being independent of t)

$$U\nabla . U = \bar{U}\nabla . \bar{U} + (\bar{U}\nabla . u + u\nabla . \bar{U}) + u\nabla . u.$$

The x-component of $(\bar{U}\nabla . u + u\nabla . \bar{U})$ is

$$\left(\bar{U}_x \frac{\partial}{\partial x} + \bar{U}_y \frac{\partial}{\partial y} + \bar{U}_z \frac{\partial}{\partial z}\right) u_x + \left(u_x \frac{\partial}{\partial x} + u_y \frac{\partial}{\partial y} + u_z \frac{\partial}{\partial z}\right) \bar{U}_x$$

$$= \frac{\partial}{\partial x}(\bar{U}_x u_x) + \left(\bar{U}_y \frac{\partial u_x}{\partial y} + u_y \frac{\partial \bar{U}_x}{\partial y}\right) + \left(\bar{U}_z \frac{\partial u_x}{\partial z} + u_z \frac{\partial \bar{U}_x}{\partial z}\right),$$

the result which is also obtained by choosing components first and afterwards effecting subdivision. We may therefore consider the method of harmonic subdivision as justified when used without further comment in § 5 at (5.1).

A few preliminary considerations will now suffice to pave the way for the analysis.

(2.2) *Notation.* In the case of the treatment applicable to *individual* electrons (2.4) the notation used is quite straightforward. The use of fluxional notation to indicate differentiation following the motion of the electron is quite usual and rather more satisfactory than the use of total differential notation d/dt, since the differentiation is, in fact, partial,

$$\dot{x} = \frac{\partial x}{\partial t}, \quad \dot{y} = \frac{\partial y}{\partial t}, \quad \dot{z} = \frac{\partial z}{\partial t},$$

$$\frac{\partial}{\partial t} \equiv \left.\frac{\partial}{\partial t}\right|_{t_0},$$

t_0 where t_0 (which $< t$) represents the instant at which the electron in question left
the origin. Thus only in the time-steady state can $\dot{x}$, $\dot{y}$, $\dot{z}$ be regarded as total
differentials.

In the *general* case, the coordinate system used is

$$(x, y, z, t),$$

x, y, z where x, y, z are now independent variables, and $\partial/\partial t$ denotes partial differentiation
with respect to time at constant (x, y, z). The velocity in the general case is repre-
U_x, U_y, U_z sented by U having components U_x, U_y, U_z parallel to the axes. At a later stage,
when U_x comes in for special consideration, the suffix x is omitted. This does not
o lead to confusion and permits of the use of the suffix o to denote forward velocity
of emission. The various components of U in the harmonic sense are represented
$\bar{U}, u, u'$ by subdivision of U into components of zero order $\bar{U}$, first order u, second order u'
and so on.

(2.3) *No restriction to small amplitudes.* The values of the harmonic components
J of U are determined by assigning to the total current J the wave form

$\bar{J}, j_1, j_2, p$
$$J = \bar{J} + j_1 \sin pt + j_2 \sin^2 pt + \dots,$$

the number of available equations being equal to the number of harmonics it is desired to consider. Thus, at any selected value of x

f $$\bar{U} = f(\bar{J}),$$

g $$u = g(\bar{U}, \bar{J}, j, t),$$

h $$u' = h(\bar{U}, \bar{J}, u_1, j_1, j_2, t),$$

and so on. In general it is only necessary to determine the functions f and g. The function h was determined for planes and cylinders in the absence of a magnetic field, but the calculation is highly laborious. In cases where we are investigating the ability of a valve to give rise to oscillations it may safely be assumed that the tube and associated circuit will distinguish between harmonics and fundamental, so that we should always be able to pick out the latter. In most cases, therefore, determination of the properties of the fundamental component gives all the information necessary to the problem, and it is to be noted that with the inclusion of initial velocities and accelerations the restriction $u \ll U$ disappears, the solution obtained being entirely independent of this assumption except when applied at points sufficiently close to the anode, for the electron current to be cut off during part of the cycle. This only occurs for magnetic fields in the neighbourhood of the critical value.

The above applies equally to the solution for individual electrons, but in this case the solution is inadequate for other reasons; see (2.6). The treatment outlined in (2.4) is given chiefly in order to facilitate understanding of the main problem.

(2.4) *Equations and solution for individual electrons of the space charge.* Consider the motion of charged particles under the combined influence of a uniform magnetic field H and an electric field X which will be non-uniform in general. Let the axis H, X of z be taken along H and that of x along X. If we choose a right-handed system (x, y, z), the equations of motion of any selected particle are

$$\left.\begin{aligned} m\ddot{x} &= eX + eH\dot{y} \\ m\ddot{y} &= -eH\dot{x} \\ m\ddot{z} &= 0 \end{aligned}\right\} \qquad \text{......(1)},$$

where m and e are the mass and negative charge of an electron. m, e

The four cases to be considered are as follows: (*a*) X uniform and constant; (*b*) X uniform, but varying in time; (*c*) X non-uniform, but independent of time at a given value of x; (*d*) X non-uniform and varying with time.

The first two cases are dealt with in (2.41), the second two in (2.42).

(2.41) *Space-charge forces negligible.* (*a*) *X uniform and constant.* The integrals of (1) are as follows:

$$\dot{x} = \dot{x}_0 \cos \omega t + \dot{y}_0 \sin \omega t + \frac{X}{H} \sin \omega t \qquad \text{......(2)},$$

$$\dot{y} = -\dot{x}_0 \sin \omega t + \dot{y}_0 \cos \omega t - \frac{X}{H}(1 - \cos \omega t) \qquad \text{......(3)},$$

$$\dot{z} = 0,$$

in which the suffix 0 corresponds to initial velocity components, and ω is written $0, \omega$ for $(e/m)\,H$. The quantity ω represents the natural angular frequency of rotation of electrons about the lines of magnetic force, and will require such frequent discussion that for the future we shall refer to ω as the *field frequency*, for want of a better term. In this and in other connections the word "frequency" will be freely used where "angular frequency" is to be understood.

A further integration gives the path of the particles:

$$x=\frac{\dot{x}_0}{\omega}\sin\omega t+\frac{1}{\omega}\left(\dot{y}_0+\frac{X}{H}\right)(1-\cos\omega t) \qquad \ldots\ldots(4),$$

$$y=\frac{\dot{x}_0}{\omega}\cos\omega t-\frac{\dot{y}_0}{\omega}\sin\omega t-\frac{X}{\omega H}(\omega t-\sin\omega t) \qquad \ldots\ldots(5),$$

$$z=\dot{z}_0 t.$$

We note from (4) the effect of tangentially emitted electrons on the forward path (that in the x direction). It can be shown that the maximum value of x is affected more by the $\dot{y}_0$ term than by the $\dot{x}_0$ term.

The x component of acceleration at the origin is given by

$$\ddot{x}_0=\omega\left(\dot{y}_0+\frac{X}{H}\right)=\omega\dot{y}_0+\frac{e}{m}X \qquad \ldots\ldots(6).$$

By equation (3) there is a mean transverse velocity of drift given by $-X/H$. This was referred to by Chapman as *the drift velocity of the free charge**. Chapman also drew attention to the fact that the mean value of $\dot{x}$ was zero despite the electric field being in the direction of x. This conclusion corresponds, of course, to perfect freedom of motion of the particles. In the case of a valve the anode may collect electrons so that the cycloidal path is interrupted, and in this case there will be a resultant motion of electrons in the x-direction.

By direct integration of the second equation of (1) we see that

$$\dot{y}=\dot{y}_0-\omega x \qquad \ldots\ldots(7).$$

This equation shows how the drift velocity varies with distance from the origin. It also shows that if we take up a position of observation at any selected value of x and watch charges moving in the yz plane, the average effect observed will be that of a constant drift velocity. It is to be noted that equation (7) holds good in all the cases (*a*) to (*d*) under consideration.

(*b*) *X uniform, but varying in time.* Let

$\bar{X}$, X_1

$$X=\bar{X}-X_1\cos pt \qquad \ldots\ldots(8),$$

in which p will be referred to as *the impressed frequency*. Thus, in the concrete case of the magnetron oscillator, p would depend on the natural frequency of the associated circuit. On the other hand, p is also to be thought of as the frequency at which the system will oscillate, if it oscillates at all. We thus adopt the method of impressing a disturbance of frequency p on the system and studying the reactions of the system to that disturbance. The forward-velocity component may be obtained by solving for $\dot{x}$ by means of the equation

$$\ddot{\dot{x}}+\omega^2\dot{x}=\frac{e}{m}pX_1\sin pt \qquad \ldots\ldots(9).$$

The solution obtained is

$$\dot{x}=\dot{x}_0\cos\omega(t-t_0)+\frac{\ddot{x}_0}{\omega}\sin\omega(t-t_0)+\frac{e}{m}pX_1\frac{1}{p^2-\omega^2}$$
$$\times\left[-\sin pt+\sin pt_0\cos\omega(t-t_0)+\frac{p}{\omega}\cos pt_0\sin\omega(t-t_0)\right] \qquad \ldots\ldots(10),$$

---

* *Proc. roy. Soc.* A, **122**, 378 (1929).

in which t_0 represents the instant at which the electron in question left the origin. t_0

It must be noted that $\ddot{x}_0$ is now a function of t_0 given by substituting (8) in (1) and putting $t=t_0$

$$\ddot{x}_0=\omega\dot{y}_0+\frac{e}{m}(\bar{X}-X_1\cos pt_0).$$

(2.42) *Space-charge forces not negligible.* (*c*) *X non-uniform, but independent of time at a given value of x.*

The non-uniformity of the field may be expressed by the Maxwellian equation

$$\dot{X}=4\pi c^2 J \qquad \text{......(11)},$$

where J is Maxwell's total current. Since J will here be independent of time, the effect of space-charge is to give rise to a field X such that J

$$X=4\pi c^2\bar{J}.t+X_0,$$

$\bar{J}$

in which t_0 is taken as zero at the origin to indicate the time steady state.

Instead of equation (8) we require the equation

$$\dddot{x}+\omega^2\dot{x}=4\pi\bar{J} \qquad \text{......(12)},$$

the steady-state solution of which is

$$\dot{x}=\dot{x}_0\cos\omega t+\frac{\ddot{x}}{\omega}\sin\omega t+4\pi\frac{e}{m}c^2\frac{\bar{J}}{\omega^2}(1-\cos\omega t) \qquad \text{......(13)},$$

in which $\ddot{x}_0$ is given by equation (6).

If we put $\bar{J}$ equal to 0 equation (13) reduces to equation (2), as would be expected, since if there is no space current the number of electrons will be insufficient to give rise to a non-uniform field. It is important to note that $\bar{J}$ represents the contribution of all electrons to the space current: equation (11) is fundamental and applies exactly whether or no a velocity-distribution exists.

(*d*) *X non-uniform and varying with time.* This is the general case. Writing

$$J=\bar{J}+j_1\sin pt$$

in equation (12), we obtain the following solution:

$$\dot{x}=\dot{x}_0\cos\omega(t-t_0)+\frac{\ddot{x}_0}{\omega}\sin\omega(t-t_0)+4\pi\frac{e}{m}c^2\frac{\bar{J}}{\omega^2}\{1-\cos\omega(t-t_0)\}$$
$$+\frac{4\pi\frac{e}{m}c^2 j_1}{p^2-\omega^2}\left[-\sin pt+\sin pt_0\cos\omega(t-t_0)+\frac{p}{\omega}\cos pt_0\sin\omega(t-t_0)\right] \quad \text{...(14)}.$$

The value of $\ddot{x}_0$ is related to that of the field at the origin. We have, by integration of (11) with respect to t over the interval $(t-t_0)$,

$$X-X_0=4\pi c^2\int_{t_0}^{t}J\,dt=4\pi c^2\left[\bar{J}(t-t_0)+\frac{j_1}{p}(\cos pt_0-\cos pt)\right] \quad \text{...(15)}.$$

The right-hand side vanishes when $t=t_0$, and we see that X_0 remains indeterminate. This, however, would be expected since it is not sufficient to know the space current, but we require also to know the total emission. If n_0 be the number of electrons emitted with forward components of velocity lying between $\dot{x}_0$ and $\dot{x}_0 + d\dot{x}_0$, and if J_0 denote total emission plus displacement current, both reckoned per cm.^{2} of cathode surface,

$$J_0 = \int_0^\infty n_0 e\, d\dot{x}_0 + \frac{1}{4\pi c^2} \frac{\partial X_0}{\partial t_0}\bigg|_{\text{cathode}} \qquad \ldots\ldots(16).$$

Now the Schottky* effect results in a slight variation of total emission with field-strength, but the alternating field itself may be considered insufficient to be effective in this respect. Again, the total emission is subject to random variations known as "Schroteffekt"†; these variations, which amount to a very small fraction of the total emission, could if desired be represented by regarding J_0 as modulated by the current to which they give rise. For present purposes we may neglect both Schottky and Schroteffekts. Then n_0 will be independent of t_0, i.e. the space charge *immediately* in front of the cathode is independent of the alternating field at the cathode, a consideration which also follows from the inability of the electrons to respond to periodic forces immediately on emerging from the cathode. Thus n_0 and $\dot{x}_0$ will be independent of t_0. We may then split equation (16) into two equations

$$\bar{J}_0 = \int_0^\infty n_0 e\, d\dot{x}_0$$

and

$$j_1 \sin pt_0 = \frac{1}{4\pi c^2} \frac{\partial X_0}{\partial t_0}\bigg|_{\text{cathode}} \qquad \ldots\ldots(17).$$

On integrating equation (17) with respect to t_0, and subject to the condition that $X_0 = \bar{X}_0$ when $t_0 = 0$,

$$X_0 = \bar{X}_0 + \frac{4\pi c^2 j_1}{p}(1 - \cos pt_0) \qquad \ldots\ldots(18).$$

The acceleration at the cathode is given by

$$\ddot{x}_0 = \omega \dot{y}_0 + 4\pi \frac{e}{m} c^2 j_1 (1 - \cos pt_0) \frac{1}{p} \qquad \ldots\ldots(19).$$

The electric field at any point of the system is given by equation (15), which with the aid of equation (18) becomes

$$X = \bar{X}_0 + 4\pi c^2 \left[\bar{J}\,(t - t_0) + \frac{j_1}{p}(1 - \cos pt)\right] \qquad \ldots\ldots(20).$$

It will be seen that the field frequency ω does not enter into the expression, a result that appears to conflict with the potential-distribution obtained in § 4, which does depend on ω. The apparent paradox is explained by the alteration of $(t - t_0)$. The electric field as given by equation (20) is thus to be regarded as modified by a magnetic field through modification of the transit time.

(2.5) *Examination of the solution in particular cases.* When the field-frequency ω becomes equal to the impressed frequency p, casual inspection of the solutions (10) and (14) suggests a resonance phenomenon similar to that corresponding to dispersion phenomena in general. With due care in the taking of limits, however, equation (14) is seen to remain finite in the case where $\omega = p$, and to assume the value

---

* Schottky, *Ann. Phys.*, Lpz., **44**, 1011 (1914).

† *Ibid.* **57**, 541 (1918); **68**, 157 (1922).

$$\dot{x} = \dot{x}_0 \cos p\,(t-t_0) + \frac{\ddot{x}_0}{p} \sin p\,(t-t_0) + 4\pi \frac{e}{m} c^2 \bar{J} \frac{1}{p^2} \{1 - \cos p\,(t-t_0)\}$$
$$+ \frac{2\pi \frac{e}{m} c^2 j_1}{p^2} [\cos pt_0 \sin \omega\,(t-t_0) - p\,(t-t_0) \cos pt] \quad \text{......(21)}.$$

It will appear later that a form of resonance can occur when $\omega = p$. For the time being we note that equation (21) shows no signs of a marked resonance effect.

Another case of interest is that of small magnetic field, $\omega \ll p$. In this case equation (14) reduces to

$$\dot{x} = \dot{x}_0 + \ddot{x}_0\,(t-t_0) + 2\pi \frac{e}{m} c^2 \bar{J}\,(t-t_0)$$
$$+ \frac{4\pi \frac{e}{m} c^2 j_1}{p^2} [-\sin pt + \sin pt_0 + p\,(t-t_0) \cos pt_0] \quad \text{......(22)}.$$

The cases $\omega = p$, $\omega \ll p$ may be compared by considering the square-bracket terms of equations (21) and (22) in the particular case $t_0 = 0$. Writing

$$\beta = 4\pi \frac{e}{m} c^2 j_1,$$

we have to compare the expressions

$$\frac{\beta}{2p^2} (\sin pt - pt \cos pt)$$

and

$$\frac{\beta}{p^2} (pt - \sin pt),$$

in which t is now to be regarded as the instantaneous value of the transit time. For small values of pt we have in each case

$$\tfrac{1}{6} \beta\, pt^3,$$

i.e. the expressions are identical near the cathode. This result suggests that the magnetic field is without effect on the velocity of electrons near the cathode. This aspect of the solution is dealt with in § 4 at (4.5). For large values of t we obtain, if $pt \gg 1$,

$$\frac{\beta}{2p^2} (-pt \cos pt)$$

and

$$\frac{\beta}{p^2} (pt),$$

a result which shows that the amplitude of motion in the case of a small magnetic field may be double that obtaining in the case of a field-frequency equal to the impressed frequency.

(2.6) *The inadequacy of particle dynamics.* From the mathematical point of view equation (14) represents the solution of the whole problem, for in order to embrace all the electrons it would merely be necessary to sum the solution over all values of the emission-velocity components $\dot{x}_0$ and $\dot{y}_0$. Furthermore, if it is desired to express the solution in terms of (x, y, z, t) as independent variables, there is a technique of transformation for this outlined in appendix 1. Essentially, then, the problem is solved. When, however, we attempt to apply the technique of appendix 1 to the case under consideration, we find that it breaks down on grounds of intractability. It is therefore impracticable to convert from one coordinate system to the other except in special cases. It follows that if a solution is required in terms of (x, y, z, t) as independent variables, then it must be obtained by working throughout in these variables. It will be seen that the only disadvantage of working in terms of the variables (x, y, z, t), or equivalent variables, disappears when once a method has been found for handling the differential equations.

The inadequacy of the solution (14) as it stands arises from the existence of a variation time arising from fluctuations in the transit time $(t-t_0)$. What is required for vacuum-tube purposes is a solution which has a readily ascertainable value at a given value of (x, y, z).

§ 3. A GENERAL THEORY OF SPACE-CHARGE PHENOMENA IN A HIGH VACUUM

(3.1) *General equations for electron-motion on classical electromagnetic theory.* The medium in which the electrons are moving possesses unit permeability and dielectric constant prior to the introduction of electrons. Because of the presence of electrons some modification in the dielectric constant will result. On the concepts of the electron theory, however, such modification leaves the dielectric constant of the space between the electrons unaltered, so that in the equations which follow we shall put $\kappa=\mu=1$ everywhere. The effective change in dielectric constant comes about indirectly through the non-uniformity of the electric field resulting from the distribution of electrons, and since the effect of this non-uniformity is included in the first of the equations below it would be definitely incorrect to include a quantity κ to represent the altered dielectric constant, for then the effect would be included twice over.

In the six equations which follow all quantities are expressed in electromagnetic units.

V The equation satisfied by the electric potential V is

$$\nabla^2 V - \frac{1}{c^2}\frac{\partial^2 V}{\partial t^2} = -4\pi c^2 \rho \qquad \text{......(23)},$$

ρ where ρ is the density of distribution of electrons at point (x, y, z).

Maxwell's total current density J is given by

$$J = \rho U + \frac{1}{4\pi c^2}\frac{\partial E}{\partial t} \qquad \text{......(24)},$$

U, E where U is the velocity and E the electric field.

The equation of motion is

$$\frac{\partial U}{\partial t} + U\nabla . U = \frac{e}{m}\{E - [U, H]\} \qquad \text{......(25)},$$

H where H is the magnetic field.

We have also the two Maxwellian equations

$$4\pi J = \operatorname{curl} H \qquad \text{......(26)},$$

$$-\frac{\partial H}{\partial t} = \operatorname{curl} E \qquad \text{......(27)},$$

while the equation of continuity takes the simple form

$$\operatorname{div} J = 0 \qquad \ldots\ldots(28).$$

Equations (23) to (28) are necessary equations for the determination of the six quantities V, ρ, J, U, E and H. It does not follow that six equations are mathematically sufficient, in view of the existence of partial differentials.

In § 2 at (2.1) it was pointed out that if the magnetic vectors of two electromagnetic waves meeting in space cancel one another the general problem would be considerably simplified. Now in equations (25) and (26) the magnetic vectors are included in H, and if we drop the magnetic vectors H will reduce to a known quantity, that representing the joint effect of the magnetic field set up by the electron current and that corresponding to the applied magnetic field. In order, however, to appreciate to the full the extent of the simplification consider the equation

$$E = -\frac{\partial \Omega}{\partial t} - \operatorname{grad} V \qquad \ldots\ldots(29),$$

where Ω represents the magnetic vector potential. Ω

If now Ω is equated to zero, equation (29) becomes

$$E = -\operatorname{grad} V \qquad \ldots\ldots(30).$$

We may then eliminate E from equations (24), (25) and (27), thus reducing the number of unknowns to five, while still retaining six equations intact.

(3.2) *Case in which magnetic field is constant.* If the applied magnetic field be invariant in time, the only alternating magnetic field present will be that corresponding to secondary magnetic effects, and these are negligible for an electron-velocity low compared with c. We may then equate $\partial H/\partial t$ to zero in equation (27), obtaining with equation (30) the known result for scalar quantities,

$$\operatorname{curl} (\operatorname{grad} V) = 0.$$

Secondary electric effects are negligible for the same reason, which means that the electric potential in equation (23) need not be considered as a retarded potential, the term $c^{-2}\, \partial^2 V/\partial t^2$ may be omitted, and we are left with Poisson's equation. The last term of equation (24) also includes secondary electric effects, but is itself by no means negligible in the case where the displacement current between a pair of electrodes is contemplated. We shall not require equation (26), since the effect on the electrons themselves of the magnetic field produced by their own current is a secondary magnetic effect and as such is being considered negligible. Our general equations are now

$$\nabla^2 V = -4\pi c^2 . \rho \qquad \ldots\ldots(31),$$

$$J = \rho U - \frac{1}{4\pi c^2} \frac{\partial}{\partial t} (\operatorname{grad} V) \qquad \ldots\ldots(32),$$

$$\frac{\partial U}{\partial t} + U \nabla . U = -\frac{e}{m} \{\operatorname{grad} V + [U, H]\} \qquad \ldots\ldots(33),$$

$$\operatorname{div} J = 0 \qquad \ldots\ldots(34).$$

In our application of the above equations we shall restrict ourselves to the cases where H is known everywhere. There are thus only four unknowns—V, ρ, J and U.

(3·3) *Application to plane diode with uniform magnetic field parallel to plates.* Choosing the electric field parallel to the axis of x and normal to the plates, and the magnetic field along the axis of z, we note that terms in $\partial U_x/\partial y$, $\partial U_x/\partial z$, $\partial U_y/\partial y$, $\partial U_y/\partial z$ may be omitted on account of considerations of symmetry, while U_z is unaffected by either field and $\partial U_y/\partial t$ is zero in the absence of periodic electric forces in the yz plane. Thus only two of the nine* terms of $U\nabla . U$ differ from zero, and the equations required are simply

$$\frac{\partial^2 V}{\partial x^2} = -4\pi c^2 . \rho \qquad \ldots\ldots(35),$$

$$J_x = \rho U_x - \frac{1}{4\pi c^2}\frac{\partial^2 V}{\partial x \partial t} \qquad \ldots\ldots(36),$$

$$\frac{\partial U_x}{\partial t} + U_x \frac{\partial U_x}{\partial x} = -\frac{e}{m}\frac{\partial V}{\partial x} + \omega U_y \qquad \ldots\ldots(37),$$

$$U_x \frac{\partial U_y}{\partial x} = -\omega U_x \qquad \ldots\ldots(38),$$

$$\frac{\partial J_x}{\partial x} = 0 \qquad \ldots\ldots(39).$$

(3·4) *Derivation of primary equation.* Equation (38) gives by direct integration, after division throughout by U_x

$$U_y = -\omega x \qquad \ldots\ldots(40).$$

The value of U_y at $x=0$ is zero since the y component of emission velocity vanishes when averaged over all electrons. With the help of equation (40), we can eliminate U_y so that equation (37) becomes

$$\frac{\partial U_x}{\partial t} + U_x \frac{\partial U_x}{\partial x} = -\frac{e}{m}\frac{\partial V}{\partial x} - \omega^2 x \qquad \ldots\ldots(37a).$$

We differentiate this equation partially first with respect to t

$$\frac{\partial}{\partial t}\left(\frac{\partial U_x}{\partial t} + U_x \frac{\partial U_x}{\partial x}\right) = -\frac{e}{m}\frac{\partial^2 V}{\partial x \partial t} \qquad \ldots\ldots(41),$$

and then with respect to x instead of t

$$\frac{\partial}{\partial x}\left(\frac{\partial U_x}{\partial t} + U_x \frac{\partial U_x}{\partial x}\right) = -\frac{e}{m}\frac{\partial^2 V}{\partial x^2} - \omega^2 \qquad \ldots\ldots(42).$$

Multiplying equation (42) by U_x and adding to equation (41), we obtain

$$\left(\frac{\partial}{\partial t} + U_x \frac{\partial}{\partial x}\right)\left(\frac{\partial U_x}{\partial t} + U_x \frac{\partial U_x}{\partial x}\right) = -\frac{e}{m}\left(\frac{\partial}{\partial t} + U_x \frac{\partial}{\partial x}\right)\frac{\partial V}{\partial x} - \omega^2 U_x$$

or

$$\left(\frac{\partial}{\partial t} + U_x \frac{\partial}{\partial x}\right)^2 U_x = 4\pi \frac{e}{m} c^2 J_x - \omega^2 U_x \qquad \ldots\ldots(43).$$

Equation (43) is the starting point for further calculations and will be referred to as the *primary equation*. It is to be noted that the primary equation contains only the unknowns U_x and J_x and we are able to impress a known value of J_x measurable outside the system. The unknowns U_x, ρ and V will all be expressible in terms of J_x once the solution for U_x is obtained. We shall thus have a relation between V and J_x, both of which are measurable* in the external circuit. The advantage of working with a known impressed current rather than a known impressed potential arises from the following consideration: when the current is known in the external circuit it is also known at all points between the plates, in view of equation (39). Every attempt to obtain a solution by working with V and J has so far proved discouraging. It is especially convenient to work with U when a magnetic field is present in view of the last term of equation (43).

---

* For the x component of $U\nabla . U$ see (2.1).

§ 4. SOLUTION FOR THE STEADY STATE

(4.1) *Average velocity and acceleration.* We shall require equation (43) in which, for the steady state,

$$\frac{\partial}{\partial t}=0, \quad \frac{\partial}{\partial x}=\frac{d}{dx}, \quad U_x=\bar{U}, \quad J_x=\bar{J}.$$

Writing $$4\pi\frac{e}{m}c^2\bar{J}=C,$$ C

we may henceforth think of C as the *steady current*, the constant $4\pi\,(e/m)\,c^2$ being understood. Equation (43) reduces to

$$\bar{U}\frac{d}{dx}\left(\bar{U}\frac{d\bar{U}}{dx}\right)=C-\omega^2\bar{U} \qquad \text{......(44)}.$$

If T is the transit time in the steady state, we have T

$$\bar{U}\frac{d}{dx}=\frac{d}{dT}.$$

We now introduce the variable θ, which will be used throughout and is defined by θ

$$\theta=\omega T \qquad \text{......(45)}.$$

Since θ/ω represents the transit time to the plane x, we may think of θ as "x at a given ω." We may also think of θ as "ω at a given x."

Equation (44) becomes

$$\frac{d^2\bar{U}}{d\theta^2}+\bar{U}=\frac{C}{\omega^2}.$$

The solution of which is of the form

$$\bar{U}=\frac{C}{\omega^2}+A_1\cos\theta+A_2\sin\theta \qquad \text{......(46)},$$

$$\frac{d\bar{U}}{d\theta}=A_2\cos\theta-A_1\sin\theta \qquad \text{......(47)}.$$

---

[*] The question whether acceptable accuracy can be achieved with existing technique depends upon a number of considerations, such as frequency, amplitude, and ratio of alternating to direct current or voltage. In general the measurements would present difficulties.

A_1, A_2 The arbitrary constants A_1 and A_2 are completely determined by assigning values to $\overline{U}$ and $d\overline{U}/d\theta$ at the cathode, which is taken to coincide with the plane $x=0$. We only have to consider the components normal to the cathode. As the y and z components of velocity vanish at the cathode when averaged over all electrons with y and z components ranging between $-\infty$ and $+\infty$, the x component of initial velocity is also the initial velocity itself in the coordinate system now in use. This is nearly, but not quite, true of the initial acceleration as will appear immediately.

Accordingly the initial velocity is given by

$$(U_0, 0, 0) \qquad \qquad \ldots\ldots(48).$$

The initial acceleration has the components

$$\left(\frac{e}{m}X_0, -\omega U_0, 0\right) \qquad \qquad \ldots\ldots(49),$$

X_0 where X_0 is the electric field at the cathode. The y component of initial acceleration, which arises from the x component of velocity in view of equation (38), does not here concern us, since equations (46) and (47) refer only to the x components of

U_0 velocity and acceleration respectively. Writing $\overline{U}=U_0$, when $\theta=0$, and also

$$\frac{d\overline{U}}{d\theta}=\frac{1}{\omega}\frac{e}{m}X_0,$$

we have at the cathode

$$U_0=\frac{C}{\omega^2}+A_1 \qquad \qquad \ldots\ldots(46a),$$

$$\frac{1}{\omega}\frac{e}{m}X_0=A_2 \qquad \qquad \ldots\ldots(47a).$$

Then boundary conditions (48), (49) suffice to determine $\overline{U}$ everywhere in terms of C and θ. Equations (46a), (47a) determine the arbitrary constants A_1 and A_2. Accordingly equations (46) and (47) become

$$\overline{U}=\frac{C}{\omega^2}(1-\cos\theta)+U_0\cos\theta+\frac{1}{\omega}\frac{e}{m}X_0\sin\theta,$$

$$\frac{d\overline{U}}{d\theta}=\frac{C}{\omega^2}\sin\theta-U_0\sin\theta+\frac{1}{\omega}\frac{e}{m}X_0\cos\theta.$$

n_1, n_2 For convenience we define n_1 and n_2 as follows:

$$\left.\begin{aligned} n_1&=1-\frac{\omega^2 U_0}{C}\\ n_2&=\omega\left(\frac{e}{m}X_0\right)C^{-1}\end{aligned}\right\} \qquad \qquad \ldots\ldots(50).$$

Typical values of n_1 and n_2 are 0·88 and $-0{\cdot}48$.

The quantity n_1 may be thought of as a correcting factor for finite emission velocities, and n_2 as proportional to the acceleration at the cathode. The fact that n_1 and n_2 reduce to unity and zero respectively when $\omega=0$ does not mean that the corrections disappear for zero magnetic field. The values of $\overline{U}$ and $d\overline{U}/d\theta$ in the simplified notation are as follows:

$$\overline{U}=\frac{C}{\omega^2}(1-n_1\cos\theta+n_2\sin\theta) \qquad \text{......(51)},$$

$$\frac{d\overline{U}}{d\theta}=\frac{C}{\omega^2}(n_1\sin\theta+n_2\cos\theta) \qquad \text{......(52)}.$$

If now $\omega\rightarrow 0$ we obtain, after writing $\theta=\omega T$,

$$\overline{U}=\frac{1}{2}CT^2+U_0+\frac{e}{m}X_0T \qquad \text{......(51}a\text{)},$$

$$\frac{d\overline{U}}{dT}=CT+\frac{e}{m}X_0 \qquad \text{......(52}a\text{)}.$$

The quantity
$$\frac{1}{\overline{U}}\frac{d\overline{U}}{d\theta}=\frac{n_1\sin\theta+n_2\cos\theta}{1-n_1\cos\theta+n_2\sin\theta} \qquad \text{......(53)}$$

will later be seen to be of fundamental importance.

The drift velocity is given by $U_y=-\omega x$ ......(54).

(4.2) *The average path of electrons.* In the steady state we may write

$$U_x=\frac{dx}{dT}, \qquad U_y=\frac{dy}{dT}.$$

By integration of equation (51) with respect to T, we obtain when $x=0$ and $T=0$:

$$x=\frac{C}{\omega^3}(\theta-n_1\sin\theta-n_2\cos\theta+n_2) \qquad \text{......(55)}.$$

A further integration gives, if $y=y_0$ when $\theta=0$,

$$y=y_0-\frac{C}{\omega^3}\left(\frac{\theta^2}{2}+n_1\cos\theta-n_1-n_2\sin\theta+n_2\theta\right) \qquad \text{......(56)}.$$

The full-line curve of figure 1 gives the shape of path for electrons emitted with negligible velocity from the cathode under conditions of space-charge limitation.

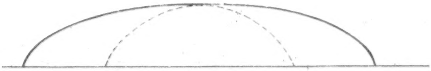


Figure 1.

The well-known cycloidal path* corresponding to zero space charge is shown dotted on the same diagram. The marked difference brought about by space charge may be understood by writing x in the form

$$x=\frac{C}{\omega^3}(\theta-n_1\sin\theta)+\frac{1}{\omega^2}\left(\frac{e}{m}X_0\right)(1-\cos\theta). \qquad \text{......(55}a\text{)}.$$

If we start with a very low cathode temperature, the anode current will be so small that the term in C is unimportant and may be omitted.

The maximum value of x is then seen to occur when $\theta=\pi$, the corresponding path being the dotted curve of figure 1. As the cathode temperature is raised the term in C begins to be of importance and $(e/m)\,X_0$ will become less on account of space-charge limitation. The path will then be intermediate between the dotted and full-line curves. If the cathode temperature be further raised to that required for space-charge saturation of the space, X_0 vanishes and the full-line curve obtains. Any further increase in cathode temperature makes $(e/m)\,X_0$ negative but leaves the path substantially unaltered. The full-line curve may thus be regarded as corresponding to practical conditions in which the space current, when magnetic field is absent, is less than the saturation value.

---

* J. J. Thomson, *Phil. Mag.* **48**, 517 (1899).

(4.3) *Critical field.* The solution for individual electrons is here of assistance as it clearly demonstrates the effect of initial velocities on the turning-point of electrons. In the steady state, equation (13) shows that the velocity of electrons with zero emission velocities is

$$\dot{x} = \frac{\ddot{x}_0}{\omega}\sin\theta + 4\pi\frac{e}{m}c^2\frac{\bar{J}}{\omega^2}(1-\cos\theta) \qquad \ldots\ldots(13a),$$

in which we have written $\omega t = \theta$ (as we are entitled to do in the steady state) so as to avoid confusion which might arise from the different meanings attached to t in the respective co-ordinate systems. Since the transit time will vary slightly over individual electrons, θ/ω should be thought of as that interval of time which has elapsed since any specified set of electrons left the cathode. In § 5 at (5.1) the conception of *transit time to a given plane* is adopted as it greatly aids the understanding of the time-variable case. The given "plane" has actually a finite thickness due to initial velocity-distribution. As shown in § 4 at (4.4) the thickness of this "plane" is of the order of 10^{-3} cm. only.

If, then, $\theta = 2\pi$, we find from equation (13a) that $\dot{x} = 0$, i.e. the electrons emitted with zero velocity come to rest. If the magnetic field be of such a value that $\theta < 2\pi$ everywhere, all electrons will reach the anode, but if $\theta > 2\pi$ some or all of the electrons will be returned to the cathode. It is convenient to define that magnetic field which makes $\theta = 2\pi$ at the anode under the given conditions as *the critical field*. This definition results in an average velocity at the anode equal to the average initial velocity of electrons. It was pointed out in § 2 at (2.41) that the excursion of the electrons was affected more by the y component than by the x component of initial velocity. This conclusion was intended to apply only to the case in which space-charge is negligible. The y component of velocity will not affect the position of the anode corresponding to critical field. Thus, by integration with respect to t, equation (13a) becomes

$$x = \frac{\ddot{x}_0}{\omega^2}(1-\cos\theta) + 4\pi\frac{e}{m}c^2\frac{\bar{J}}{\omega^3}(\theta-\sin\theta) \qquad \ldots\ldots(57).$$

Since, by equation (6),
$$\ddot{x}_0 = \omega\dot{y}_0 + \frac{e}{m}X_0,$$

we see that when $\theta = 2\pi$ the value of $\dot{y}_0$ is without effect on the value of x. Writing

$$4\pi\frac{e}{m}c^2\bar{J} = C,$$

we have for the cathode-anode separation for critical field

$$d = \frac{C}{\omega^3}2\pi \qquad \ldots\ldots(58),$$

or writing
$$\omega = \frac{e}{m} H_c$$
H_c

and rearranging
$$H_c = \frac{m}{e}\left(2\pi C d^{-1}\right)^{\frac{1}{3}} \qquad \text{......(58}a\text{)}.$$

Equation (58 *a*) gives the critical field in terms of the anode current and the anode-cathode separation.

By integration of equation (37*a*) in § 3 with respect to x we obtain in the steady state, subject to $V=0$ when $x=0$,
$$U_x^2 = U_0^2 - 2\frac{e}{m} V - \omega^2 x^2.$$

For the anode potential corresponding to the critical condition, if
$$x=d, \quad U_x = U_0, \quad \omega = eH_c/m,$$
$$V_c = (-e/2m)\, H_c^2 d^2 \qquad \text{......(59)}.$$

Eliminating H_c between equations (58*a*) and (59) we obtain the relation between space current and anode voltage corresponding to the critical condition
$$J_c = \frac{(2e/-m)^{\frac{1}{2}} V_c^{\frac{3}{2}}}{(2\pi C d)^2} \qquad \text{......(60)},$$
which is seen to be only $9/4\pi$ times the value without magnetic field*.

Since in the critical condition even the slowest electrons are only just brought to rest, the above reduction in anode current as compared with the conditions obtaining for zero magnetic field is not attributable to returning electrons. The reduction may be regarded as due to increased space-charge limitation. This point will be considered again in (4.5).

(4.4) *Magnetic field in excess of critical value.* The y component of velocity, or drift velocity, will still be in the same direction, but the x component of velocity will be reversed in sign for those electrons which are returning to the cathode. If a fraction κ of the outgoing electrons return to the cathode, then the space-charge κ density is increased in the ratio $(1+\kappa)$ provided we are only considering the steady state, in which electrons move back with a velocity which, at a given plane, is equal and opposite to the outgoing velocity.

Let J_c correspond to the space current flowing when the magnetic field is of J_c value just insufficient to return any electrons to the cathode. If ρ_c represent the ρ_c space-charge density under the above conditions, we have under the new conditions a nett space-charge density ρ equal to $\rho_c(1+\kappa)$. ρ

Let U_x, U_y be the velocity components under the new conditions; then for U_x, U_y outgoing electrons we have the current J_1, where
$$J_1 = \rho_c U_x.$$
J_1

For returning electrons we have the current J_2, where
$$J_2 = -\kappa \rho_c U_x.$$
J_2

---

* Langmuir, *Phys. Rev.* **2**, 450 (1913).

J

Then the resultant current J in the x direction is given by

$$J = J_1 + J_2 = \rho_c (1+\kappa)\, U_x = \rho U_x \frac{1-\kappa}{1+\kappa} \qquad \text{......(61).}$$

The appearance of $(1+\kappa)$ in the denominator signifies additional space-charge limitation of current.

Now
$$\frac{\partial^2 V}{\partial x^2} = -4\pi c^2 \rho,$$

while
$$U_x \frac{\partial U_x}{\partial x} = -\frac{e}{m}\frac{\partial V}{\partial x} + \omega U_y$$

and
$$U_x \frac{\partial U_y}{\partial x} = -\omega U_x \qquad \text{......(62);}$$

after writing from equation (62), by integration,

$$U_y = -\omega x \qquad \text{......(63),}$$

we have
$$U_x \frac{\partial U_x}{\partial x} = -\frac{e}{m}\frac{\partial V}{\partial x} - \omega^2 x \qquad \text{......(64),}$$

whence
$$U_x \frac{\partial}{\partial x}\left(U_x \frac{\partial U_x}{\partial x}\right) = 4\pi \frac{e}{m} c^2 J \frac{1+\kappa}{1-\kappa} - \omega^2 U_x \qquad \text{......(65),}$$

where
$$\left.\begin{aligned} J_c &= J \frac{1-\kappa}{1+\kappa} \\ C_c &= 4\pi \frac{e}{m} c^2 J_c \end{aligned}\right\} \qquad \text{......(66).}$$

The solution is substantially as before, e.g.

$$U_x = \frac{C_c}{\omega^2}(1 - n_1 \cos\theta + n_2 \sin\theta).$$

C_o

We can now determine the value of U_x for magnetic fields in excess of the critical value, since even if all the electrons are returned to the cathode, C_c is to be interpreted as the one-way current corresponding to the condition $\kappa = 0$. The value of U_x will be affected by the larger value of ω, so that at any point between the plates the outgoing velocity is lessened on this account, as also is the potential. The velocity of the returning electrons is given by equation (67), but with the sign of C_c reversed.

As in (4.3) electrons emitted from the cathode with zero velocity come to rest when $\theta = 2\pi$. In the case now under consideration such electrons reverse their direction when $\theta = 2\pi$, and will be travelling back towards the cathode for values of θ given by

$$2\pi < \theta < 4\pi.$$

The electrons of average emission velocity reverse their direction when

$$1 - n_1 \cos\theta + n_2 \sin\theta = 0 \qquad \text{......(67),}$$

which corresponds very nearly to

$$\theta = 2\pi + \frac{1-n_1}{-n_2} \qquad \text{......(68).}$$

Such electrons do not arrive back at the cathode until

$$\theta = 4\pi + \frac{1-n_1}{-n_2} \qquad \text{......(69).}$$

For a retarding field at the cathode surface n_2 will be negative. Giving n_1 and n_2 their values from equation (50), we find that

$$\frac{1-n_1}{-n_2} = \frac{\omega U_0}{-\frac{e}{m} X_0} \qquad \text{......(70).}$$

For a magnetic field of 100 G.* and a retarding field† at the cathode of 100 V./cm.

$$\left.\begin{aligned} \omega &= 1{\cdot}77 \times 10^9 \text{ sec.}^{-1} \\ \frac{e}{m} X_0 &= -1{\cdot}77 \times 10^{17} \text{ cm./sec.}^2 \end{aligned}\right\} \qquad \text{......(71).}$$

If we take 2×10^7 cm./sec. as the mean emission velocity, equation (69) becomes

$$\theta = 2\pi + \tfrac{1}{5} \qquad \text{......(72).}$$

This corresponds to a change of some 3 per cent in θ. The maximum distance travelled in the x direction by electrons emitted with average velocity is given by $\bar{x}$, where $\bar{x}$

$$\bar{x} = \frac{C_c}{\omega^3} (\theta + n_2 - n_2 \cos\theta - n_1 \sin\theta)$$

in which

$$\theta \doteqdot 2\pi + \frac{1-n_1}{-n_2}.$$

Inserting the above value of θ we obtain after reduction

$$\bar{x} = \frac{2\pi C_c}{\omega^3} + \frac{U_0^2}{2(-e/m) X_0}.$$

With the values of U_0 and $(e/m) X_0$ given under equation (71), the value of the last term is only about 10^{-3} cm., which shows that all electrons turn back at substantially the same value of x given by $2\pi C_c/\omega^3$.

(4·5) *Potential-distribution.* The value of the potential $\bar{V}$ at any value of x may be obtained from the equation $\bar{V}$

$$2\left(\frac{-e}{m}\right)\bar{V} = \bar{U}^2 - \bar{U}_0^2 + \omega^2 x^2 \qquad \text{......(73).}$$

In this connection it is not convenient to express the velocity $\bar{U}$ in terms of x, and the most satisfactory method is to express $\bar{V}$ entirely in terms of θ and obtain the $\bar{V}$, x relation by graphical means with the aid of the θ, x relation expressed by equation (55). The $\bar{V}$, θ relation is as follows: $\bar{U}$

$$\bar{V} = \frac{mC^2}{-e\omega^4}\left[\frac{1}{2}\theta^2 - n_1\theta\sin\theta + n_1(1-\cos\theta) + n_2\{(1-n_1)\sin\theta + (\theta+n_2)(1-\cos\theta)\}\right] \qquad \text{......(74),}$$

---

* When ω is positive, H has same sign as e/m.

† X_0 is positive for a retarding field.

while from (4.2) $$x=\frac{C}{\omega^3}(\theta+n_2-n_1\sin\theta-n_2\cos\theta) \qquad \ldots\ldots(55a).$$

In the special case of zero initial velocities and accelerations we obtain the following simplified relations for $\overline{V}$ and x

$$\overline{V}=\frac{mC}{-e\omega^4}\left[\frac{1}{2}\theta^2-\theta\sin\theta+1-\cos\theta\right] \qquad \ldots\ldots(74a),$$

$$x=\frac{C}{\omega^3}[\theta-\sin\theta] \qquad \ldots\ldots(55b).$$

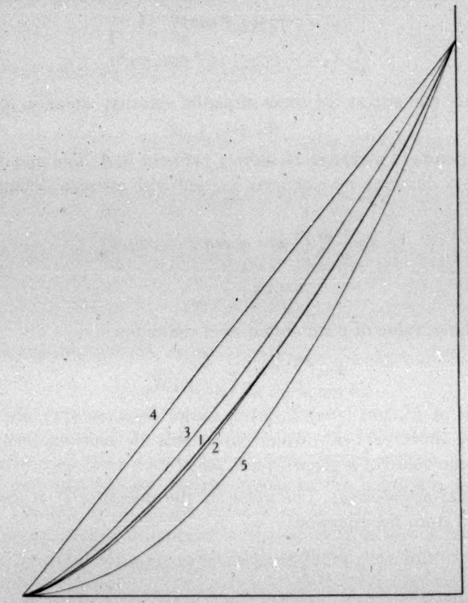


Figure 2. Effect of magnetic field on potential-distribution (planes).

Equations (74) and (55*a*) give the potential-distribution under any conditions of magnetic field, from zero to the critical value. The potential-distribution under the condition of critical field, from equations (74*a*) and (55*b*), is given by curve (1) in figure 2. For comparison with curve (1) the following curves (2) to (5) have been added to figure 2 in thin lines:

$$(2)\ V\propto x^{1\cdot5},\quad (3)\ V\propto x^{4/3},\quad (4)\ V\propto x,\quad (5)\ V\propto x^2 \qquad \ldots\ldots(75).$$

The close correspondence of curve (1) with curve (2) indicates that the potential-distribution obeys a 1·5-power law very nearly for points not too close to the cathode. Curve (3) corresponds to the potential-distribution for zero magnetic field, and curve (4) to that for negligible space charge, while curve (5) is purely hypothetical.

The potential-distribution in the presence of a magnetic field is found to differ so little from that for zero magnetic field that the analysis of previous workers* for the initial velocity-correction and potential-minimum may be used for magnetic fields up to 100 G. with an error of only 2 per cent in the minimum potential. In view of the above considerations a full treatment of the potential-distribution in the presence of a magnetic field is thought not to be warranted.

Following Richardson we write for the number dN_s of electrons emitted per unit area per second with forward components of velocity lying between $\dot{x}_0$ and $\dot{x}_0 + d\dot{x}_0$ N_s

$$dN_s = N_s \frac{m\dot{x}_0}{kT_c} e^{-m\dot{x}_0^2/2kT_c}\, d\dot{x}_0 \qquad \ldots\ldots(76),$$

in which the temperature T_c is provided with a suffix denoting "cathode," in order to distinguish it from T the transit time. T_c

If the current flowing to the anode is less than the saturation current (corresponding to N_s) this deficiency must be due to a retarding potential-gradient close to the surface of the cathode by which the more slowly moving electrons are forced back to the cathode. If the potential of the anode is positive and that of the cathode zero, there must be a surface of negative potential between cathode and anode at which the potential is a minimum. This conclusion also applies when a magnetic field is present: those electrons which fail to reach the potential-minimum when no magnetic field is present will still fail to do so when the magnetic field is applied.

If V_m represents the potential at the potential-minimum surface, J_s the saturation V_m, J_s
current and J the actual anode current, only those electrons which are emitted with J
sufficient velocity to overcome the potential-difference contribute to the anode current J.

When we take the magnetic field into account the kinetic energy of individual electrons at any point is given by

$$\tfrac{1}{2}m\dot{x}^2 = \tfrac{1}{2}m\dot{x}_0^2 - eV - \tfrac{1}{2}m\omega^2 x^2 \qquad \ldots\ldots(77).$$

The initial energy $\tfrac{1}{2}m\dot{x}_0^2$ of the electrons will enable them to overcome the potential minimum if

$$\tfrac{1}{2}m\dot{x}_0^2 \geqslant eV_m + \tfrac{1}{2}m\omega^2 x_m^2,$$

in which e and V_m are both negative, so that eV_m is positive. Electrons corresponding in number to J_s are being emitted continuously from the cathode even when the current is not saturated, but a certain fraction of them are then made to return by the retarding field existing close to the cathode.

By integrating equation (76) with respect to $\dot{x}_0$ between the limits

$$(2eV_m/m + \omega^2 x_m^2)^{\frac{1}{2}}$$

---

* Epstein, *Ber. dtsch. phys. Ges.* **21**, 85 (1919); Fry, *Phys. Rev.* **17**, 441 (1921); Langmuir, *Phys. Rev.* **21**, 419 (1923).

and ∞ and dividing by

$$N_s\left(\frac{kT_c}{2\pi m}\right),$$

we obtain the modified Boltzmann equation

$$\frac{J}{J_s}=\epsilon^{-(eV_m+\frac{1}{2}m\omega^2x_m^2)/kT_c}=\epsilon^{-eV_m'/kT_c} \qquad \ldots\ldots(78).$$

V_m'

A little consideration shows that V_m' is in fact the new value of the minimum potential. The magnetic field affects the space charge by curving the paths of the electrons. While at points very close to the cathode the direction of motion of electrons is hardly altered, the path of electrons at points further removed from the cathode is such as to increase the space charge. This in turn depresses the potential at all points, including those very near to the cathode. The term $\frac{1}{2}m\omega^2x_m^2$ may be interpreted as electric potential energy arising from the influence of the magnetic field on the space charge.

With the help of Langmuir's analysis we then find the values of V_m and x_m corresponding to negligible magnetic field, and insert these values in equation (78) to see what effect arises from the term $\frac{1}{2}m\omega^2x_m^2$.

In illustration of the procedure let us consider an example representative of present-day practice. Instead of a bright tungsten emitter let us take a chemically coated cathode operating at 900° C., so that $T_c=1173°$ K. This is the first occasion that we have had to consider the nature of the emitting surface. The theory can be applied whatever the values of A and b, the thermionic constants of Richardson*. Provided that there is no positive emission from the cathode we are free to apply the analysis to any form of emitting surface free from irregularities.

A, b

The proportion of the total emission which can safely be taken from a chemically coated cathode varies considerably with the nature of the coating, but would not generally exceed 25 per cent for a reasonable life. In the case of a typical coating operating at 1173° K. the total emission will be around 200 mA./cm.^{2} Let us then take J to be 50 mA./cm.^{2} Neglecting magnetic field and following Langmuir† we find that

$$\frac{eV_m}{kT_c}=\log_\epsilon\frac{J_s}{J}=\log_\epsilon 4=1{\cdot}3863=\eta_1.$$

ξ, η
ξ_1, η_1

Langmuir's table II connecting the variables ξ, η defined by his equations (5) and (6) may then be used to determine the value ξ_1 of ξ corresponding to η_1. We thus obtain

$$-\xi_1=1{\cdot}7950.$$

From Langmuir's equations (19), (20) and (21), if we write $x_1=0$ to correspond with the cathode, we obtain

$$\left.\begin{aligned}V_m&=-0{\cdot}14022\text{ V.}\\ x_m&=1{\cdot}7753\times10^{-3}\text{ cm.}\end{aligned}\right\} \qquad \ldots\ldots(79).$$

---

* *Emission of Electricity from Hot Bodies* (Longmans, Green & Co., 1921).

† *Loc. cit.*

From equation (78) we have

$$V_m' = V_m + \frac{m}{2e}\omega^2 x_m^2 = V_m + \frac{e}{2m}H^2 x_m^2 \qquad \text{......(80).}$$

Taking H as 100 G. and substituting for V_m and x_m from equation (79), we find that

$$V_m' = -0{\cdot}1402 - 0{\cdot}0027 = -0{\cdot}1429 \text{ V.},$$

thus confirming that the potential minimum is increased by the magnetic field by 2 per cent only when $H = 100$ G.

But for a magnetic field equal to the critical value we find by equation (58*a*) when $J = 50$ mA./cm.^{2} and $d = 0{\cdot}5$ cm.

$$H_c = 131{\cdot}60 \text{ G.}$$

Insertion of this value of H in equation (80) shows that in the example under consideration the correction for magnetic fields up to the critical value amounts to 3·5 per cent in V_m. A reduced anode current as compared with that for zero magnetic field, as given by equation (60), must receive explanation on the basis of an increased number of electrons being unable to overcome the potential-minimum. A reduction of 28·4 per cent in J seems at first sight too great to be accounted for by a 3·5 per cent increase in $(-V_m)$, but it must be borne in mind that, since $J_s = 4J$, the corresponding change in the returning current $J_s - J$ is much less, amounting to some 14 per cent only.

(4.6) *The electric field and the initial velocity.* As compared with equation (74) for the potential, we have for the electric field the refreshingly simple relation

$$X = \frac{Cm}{\omega e}(\theta + n_2) = \frac{Cm}{\omega e}\theta + X_0 \qquad \text{......(81).}$$

Equation (81), which is equivalent to equation (20) of § 2, may be confirmed by taking $-dV/d\theta$ from equation (74) and dividing by $dx/d\theta$.

When $\theta = 2\pi$ we have for the electric field

$$X = X_0 + \frac{2\pi C}{\omega}\frac{m}{e}.$$

If $\theta = 2\pi$ at the anode we obtain with the help of equations (58) and (59), writing $X = X_c$ and eliminating C/ω,

X_c

V_c

$$X_c = X_0 - 2V_c/d \qquad \text{......(82).}$$

Equation (82) gives the electric field at the anode in the critical condition. If $X_0 \leqslant 2\frac{V_c}{d}$, we note that equation (82) gives a value of $(-X_c)$ 50 per cent greater than the value $\frac{4}{3}V/d$ corresponding to zero magnetic field.

Of more importance than the electric field at the anode is the electric field at the cathode. Negative at all points to the right of the potential minimum, the electric field vanishes at the potential minimum and becomes positive at the cathode. Equation (81) shows that the value θ_m of θ corresponding to the potential minimum is given by

θ_m

$$\theta_m = -n_2.$$

T_m The transit time to the potential minimum is accordingly T_m, where

$$T_m = \frac{-n_2}{\omega} = \frac{-eX_0}{mC} \qquad \text{......(83).}$$

In order to determine T_m and X_0 let us write $\theta = \theta_m$, $n_2 = -\theta_m$ in equations (74) and (55 *a*). We then obtain

$$V_m = \frac{mC^2}{-e\omega^4}\left[\frac{1}{2}\theta_m{}^2 - \theta_m \sin\theta_m + n_1(1-\cos\theta_m)\right] \qquad \text{......(84),}$$

$$x_m = \frac{C}{\omega^3}(\theta_m \cos\theta_m - n_1 \sin\theta_m) \qquad \text{......(85).}$$

Eliminating n_1 between equations (84) and (85) we obtain

$$V_m = \frac{mC^2}{-e\omega^4}\left[\frac{1}{2}\theta_m{}^2 - \left(\theta_m + \frac{\omega^3 x_m}{C}\right)\left(\frac{1-\cos\theta_m}{\sin\theta_m}\right)\right] \qquad \text{......(86).}$$

When the magnetic field is negligible equations (85) and (86) reduce to

$$x_m = U_0 T_m - \tfrac{1}{3} C T_m{}^3 \qquad \text{......(85 a),}$$

$$V_m = -\tfrac{1}{2} X_0 x_m + \frac{m}{e}\frac{C^2}{24} T_m{}^4 \qquad \text{......(86 a).}$$

Now in (2.4) it was shown that for regions in which electrons are moving in two directions at once the value of C at any point corresponds to the outgoing current at that point. Now at the cathode we have a total outgoing current of 200 mA., but since electrons are returning to the cathode at all points between the cathode and the potential-minimum, the outgoing current continually decreases until at the potential-minimum itself the outgoing current becomes equal to the anode current, namely 50 mA. in the example of the previous section. The value of C corresponding to an outgoing current of 50 mA. is (if we take $\frac{C}{m} = 1{\cdot}766$ to correspond with Langmuir's figure)

$$C = 0{\cdot}9987 \times 10^{+27} \text{ cm./sec.}^3 \qquad \text{......(87).}$$

If V_m, x_m are regarded as known, there are still three unknowns in equations (85 *a*), (86 *a*), namely X_0, U_0 and T_m.

We must therefore use a third equation, e.g. (83), to eliminate X_0 from equation (86 *a*), when we shall be able to derive T_m. Having once determined T_m, we can then obtain X_0 and U_0 from equations (83) and (85 *a*) respectively.

Eliminating X_0 between equations (83) and (86 *a*)

$$2\frac{eV_m}{m} = Cx_m T_m + \frac{1}{12} C^2 T_m{}^4 \qquad \text{......(88).}$$

Inserting in the above equation the values of V_m, x_m and C given in equations (79) and (87), taking $-e/m$ equal to $1{\cdot}766 \times 10^7$, and e.m.u. per gm. multiplying V_m by 10^8 to bring it to e.m.u., we have

$$4{\cdot}9517 \times 10^{14} = 1{\cdot}7770 \times 10^{24}\, T_m + 8{\cdot}3112 \times 10^{52}\, T_m{}^4.$$

The solution of this equation may be found graphically and is

$$T_m = 2{\cdot}0154 \times 10^{-10} \text{ sec.} \qquad \text{......(89).}$$

With the above value of T_m equation (83) gives for the retarding field at the cathode surface

$$X_0 = 113{\cdot}96 \text{ V./cm.} \qquad \text{......(90)},$$

while from equation (85*a*), which may be written

$$U_0 = \frac{x_m}{T_m} + \frac{C T_m^2}{3},$$

we obtain for the average emission velocity of electrons constituting the anode current

$$U_0 = 2{\cdot}2330 \times 10^7 \text{ cm./sec.} \qquad \text{......(91)}.$$

The above value of U_0 exceeds by only 0·35 per cent the value U_m defined by

$$U_m = \sqrt{\left(\frac{2eV_m}{m}\right)} = 2{\cdot}2253 \times 10^7 \text{ cm./sec.} \qquad \text{......(92)},$$

so that the percentage of electrons in the anode current which have velocities greatly in excess of the minimum value required to enable them to overcome the potential-minimum must in the present case be very small. For comparison with equations (91) and (92) we have the following expression for the mean velocity of emission $\overline{U}_0$ of *all* electrons emitted from the cathode, including those which turn back to the cathode without reaching the potential-minimum

$$\overline{U}_0 = \left(\frac{\pi k T_c}{2m}\right)^{\frac{1}{2}},$$

where k is Boltzman's constant and $= 1{\cdot}372 \times 10^{-16}$ erg./deg.

This equation gives in the present example, in which $T_c = 1173°$ K., the value

$$\overline{U}_0 = 1{\cdot}6750 \times 10^7 \text{ cm./sec.}$$

The small difference between U_0 and U_m may be associated with the exponential character of the Maxwellian law of distribution of velocities, with the fact that $\overline{U}_0$ is considerably less than U_m, and with the finite probability that electrons of velocity less than U_m will succeed in penetrating the barrier*.

The data relevant to the example, and also for the case in which $J/J_s = 0{\cdot}4$ are summarized in table 1, in which columns 3 and 11 give respectively the approximate anode voltage for negligible and critical magnetic fields. As has already been pointed out, the potential-minimum and associated data will remain sensibly unaffected when a magnetic field is applied.

Table 1. $J_s = 200$ mA./cm.² $T_c = 1173°$ K. $d = 0{\cdot}5$ cm.

1	2	3	4	5	6	7	8	9	10	11
J mA./cm.²)	$\frac{J}{J_s}$	V (V.)	$-V_m$ (V.)	$10^3 x_m$ (cm.)	$10^{10} T_m$ (sec.)	X_0 (V./cm.)	$10^{-7} U_0$ (cm./sec.)	$10^{-7} U_m$ (cm./sec.)	$10^{-7} \overline{U}_0$ (cm./sec.)	V_c (V.)
50	0·25	300	0·14022	1·7753	2·0154	113·96	2·2330	2·2253	1·6750	380
80	0·4	410	0·09266	1·1925	1·3479	121·97	1·8524	1·8095	1·6750	520

---

* Quantum-mechanical considerations have generally been omitted in this paper, as they are believed to have importance only very near to the cathode. The absolute value of the potential-minimum would probably be somewhat less on wave-mechanical considerations, which seem to indicate a smaller space-charge density than classical theory.

The potential-distribution near the cathode corresponding to the above data is indicated on figure 3. The values of X_0 and U_0 given in columns 7 and 8 may be regarded as typical and will be used in future examples. It is noteworthy that X_0 comes out greater when $J/J_s=0{\cdot}4$ than when $J/J_s=0{\cdot}25$. This does not prevent a greater proportion of slowly moving electrons reaching the anode, the value of the minimum potential being the deciding factor.

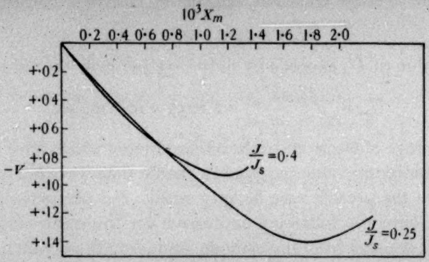


Figure 3.

(4·7) *The transit time between cathode and anode.* (4·71) *Influence of space charge.* We now come to the most important of all factors affecting the oscillations obtained in practice, whether by the magnetron method or by using a triode valve. Any error in the estimation of electron transit time might result in a completely erroneous conception of the mechanism which governs the excitation of ultra-high-frequency oscillations.

Neglecting space charge and initial velocities, Okabe* finds for the relation between transit time and critical field

$$t=\frac{\pi cm}{eH},$$

which in our present notation and units corresponds to

$$\theta_c=\pi.$$

But we have already seen (4·3) that the critical condition corresponds to

$$\theta_c=2\pi.$$

Since Okabe's result applies to both planes and cylinders in the absence of space charge effects, we cannot account for such a large difference in θ_c on the basis of electrode-geometry. Moreover, comparing Okabe's equation (9) (and putting $r'=d$ in it) with equation (59) of the present paper, we see that the relation between critical magnetic field and anode voltage is the same whether space charge is neglected or included, so that we must conclude that the difference in θ_c corresponds entirely to a difference in transit time. The 100-per-cent increase in transit time indicated by the present paper is to be regarded as a result of including the effects of space charge: it would perhaps not be altogether true to say that the increase is brought about by the space charge itself, but rather that the neglect of space charge cannot be justified except in cases where the total emission is too small to be of interest. Except for very small emission currents, equation (55) of the present paper shows that the maximum value of x is not reached by electrons until $\theta=2\pi$. In order to check this point we may try putting θ equal to π in equation (55). We then obtain

---

* *Proc. Inst. Radio Engrs*, N.Y., 17, 1, 652 (1929).

$$x=\frac{C}{\omega^3}(\pi+2n_2).$$

In all cases where the anode current is less than or just equal to the total emission, n_2 will be negative or zero and the value of x corresponding to $\theta=\pi$ will be less than or equal to *half* the value $2\pi C/\omega^3$ corresponding to $\theta=2\pi$. If the cathode-temperature is reduced considerably below that corresponding to the space-charge saturation value, then $\pi<\theta_c<2\pi$ but, as we shall see, this case is not of importance. Thus, in equation (51) for the velocity, if we write $\theta=\pi$ we find that U has the value

$$U=\frac{C}{\omega^2}(1+n_1)=\frac{2C}{\omega^2}-U_0.$$

The condition $\theta=\pi$ could only correspond to the critical condition if the anode current were of such a value that $U=U_0$, or

$$\frac{2C}{\omega^2}=2U_0,$$

giving
$$|J|=\frac{\omega^2 U_0}{4\pi|e/m|c^2}\times 10^4 \text{ mA./cm.}^2,$$

which corresponds to 6 mA./cm.² in a typical case. But we have already seen that, for $\theta=\pi$ to correspond to the critical condition, the cathode-temperature must be below the space-charge saturation value, so that we must take $J_s=J$. A total emission-density of 6 mA./cm.² is too low for oscillations to occur, and the condition $\theta=\pi$ must therefore be ruled out as corresponding to a total emission-density much smaller than the generation of oscillations in practice requires. If the anode current is only slightly below the space-charge saturation value, n_2 will be slightly positive and θ_c slightly less than 2π. While oscillations in a magnetron may sometimes occur under these conditions, the degree of space-charge saturation will always be such as to make $\theta_c=2\pi$ very nearly.

Summarizing the above, we may say that the critical condition will be given by $\theta_c=2\pi$ in all cases where the anode current is less than the total emission, and also when the anode current is equal to the total emission and the system is in the condition of space-charge saturation. A typical transit time ($J=50$ mA., $J_s/J=4$) is given by putting ω equal to $2{\cdot}33\times10^9$, which gives

$$T=2{\cdot}70\times10^{-9}\text{ sec.}^*$$

---

* From table 1 we see that T_m, which is included in T, amounts to some 7·5 per cent of T, a somewhat large percentage having regard to the fact that $x_m/d=0{\cdot}355$ per cent only. This state of affairs was predicted in *Phil. Mag.* **5**, 641 (1928). See for example figure 2 of that paper.

(4·72) ***Influence of magnetic field.*** We have next to consider to what extent the transit time is increased by the magnetic field, and for this purpose we shall require to examine our knowledge of transit times in the absence of a magnetic field.

The author first gave the formula for the transit time in the plane case* and also for the cylindrical case,† neglecting initial velocities, but including the effect of space charge. His expressions for zero magnetic field and Okabe's expressions for negligible space charge give nearly equal values for T at a given anode voltage. This suggests that the modification of transit time on account of magnetic field is just about as important as the modification on account of space charge. Megaw‡ finds by graphical solution of his equation (9) that the transit time in the cylindrical case including magnetic field is increased some 36 per cent on account of space charge if R/a is of the order of 100. McPetrie§ gives a semi-graphical method of computing transit times in the absence of a magnetic field. McPetrie's study leads to the same results as Benham's formula for cylinders and provides a rapid method for direct computations of transit time. It will appear that Megaw's estimate of 36 per cent for cylinders is much below the corresponding figure for planes, i.e. 100 per cent.

If we express our transit time in terms of the anode voltage by means of equation (59), we can compare the transit times with and without magnetic field. In the critical condition we have, if V be the anode voltage, V

$$T=\frac{2\pi}{\omega}$$

$$=\frac{2\pi d}{\sqrt{(-2Ve/m)}}.$$

If the magnetic field be removed the transit time T' corresponding to V is T' where

$$T'=\frac{3d}{\sqrt{(-2Ve/m)}}.$$

Thus $$T/T'=\tfrac{2}{3}\pi=2\cdot094.$$

If, however, we had neglected space charge we should have had for the transit time corresponding to critical field

$$T_1=\frac{\pi}{\omega}=\frac{\pi d}{\sqrt{(-2Ve/m)}},$$

in which the suffix 1 is inserted to correspond with negligible space charge. For zero magnetic field and negligible space charge the transit time is 1

$$T_1'=\frac{2d}{\sqrt{(-2Ve/m)}},$$

so that $$\frac{T_1}{T_1'}=\frac{\pi}{2}=1\cdot571.$$

---

* *Phil. Mag.* **5**, 653 (March 1928).

† *Phil. Mag.* **11**, 502, equation (46) (1931).

‡ *J. Instn elect. Engrs*, **72**, 326 (1933).

§ *Phil. Mag.* **16**, 284 (1933).

If we start with no space charge and no magnetic field, the effect of either, supposed to be separately applied, could be to increase the transit time in the ratio 3 to 2 approximately. Thus

$$\frac{T'}{T_1'}=1{\cdot}5, \qquad \frac{T_1}{T_1'}=1{\cdot}571.$$

The overall increase corresponding to space charge plus critical magnetic field as compared with zero space charge and zero magnetic field is

$$\frac{T}{T_1'}=\pi=3{\cdot}142.$$

Summarizing the information thus far available for planes and cylinders, we may construct table 2.

Table 2

	$\frac{T'}{T_1'}$	$\frac{T_1}{T_1'}$	$\frac{T}{T_1}$	$\frac{T}{T'}$	$\frac{T}{T_1'}$	Remarks
Planes	1·5	1·571	2	2·094	3·142	Dashed letters correspond to zero magnetic field, suffix 1 to zero space charge. For cylinders, R/a has been assumed to be large.
Cylinders	1·5	1·571	—	—	—	

While the transit time for cylinders including magnetic field and space charge has not yet been worked out, table 2 suggests that T/T_1 will have the value 2. Against this there is Megaw's graphically determined value $T/T_1=1{\cdot}36$*. Megaw's figure was obtained on the assumption of no change in the potential distribution due to the magnetic field, and would be somewhat too low on this account.

Regarding the equation $T/T_1=2$ as established on theoretical grounds for plane electrodes, we see that the period of the characteristic oscillations of a plane solid-anode magnetron must be equal to the time of a single transit between cathode and anode, and not to twice this transit time. The analysis of the present paper points to the conclusion that any electrons returning from anode to cathode contribute nothing to the negative resistance. The electrons which return to the cathode may in fact arrive with high-frequency energy sufficient to result in deleterious cathode bombardment. This would raise the cathode-temperature and emission. Megaw* has obtained this effect experimentally in the split-anode magnetron. The split-anode magnetron† is not considered in the present paper, nor has the effect of tilted magnetic field been included.

---

* *Loc. cit.*

† I understand from Mr Megaw that he is at present attempting an explanation of the dynatron characteristics obtained with the split-anode magnetron, taking into account the non-radial component of electric field at all points.

FOREWORD TO §§ 5 AND 6

We now come to consider in §§ 5 and 6 the ultra dynamic condition implied by the presence of rapidly alternating currents. The aim throughout these two sections is to arrive at a complete solution for the alternating-current component of electron-velocity, hereafter referred to as *alternating velocity*, in terms of the alternating current itself, and to examine the solution obtained in particular cases. The alternating current is assumed as a boundary condition; but since the value of the alternating current is specified, not only at both boundaries of the valve but also at all intermediate points, the assignment of a value to the alternating current amounts to something more than the determination of a boundary condition. As has been established in previous work*, the physical quantities potential, space-charge density and alternating velocity will all be delayed in phase with respect to the alternating current, which includes both electron and displacement current. It is, of course, possible to work out the electron current separately, and when this is done the potential is seen to lead the electron current in phase. From the practical point of view no means of distinguishing between electron and displacement current is available, since these combine at the boundaries of the valve. To be able to express the alternating velocity and potential in terms of the total alternating current is thus of paramount importance from the practical viewpoint, and in order to see whether work is done on or by the electrons all that is necessary is to multiply together the in-phase components of current and potential, and if the sign of the coefficient of $\sin^2 pt$ thus obtained is negative, power will be expended on the electrons by the alternating forces, and oscillations may result. Since the in-phase component of current is assumed positive, it is not even necessary to take the product of the in-phase components, but merely to examine the sign of the coefficient of $-\sin pt$ in the potential to see if this is ever positive. Since the quadrature component of current is zero by definition, the quadrature component of potential contributes no power either positive or negative, when averaged over a complete cycle.

The mathematics of §§ 5 and 6 require a knowledge of partial differential equations and of circular functions, but nothing else of an advanced nature. A word of warning may not be out of place to those attempting the details of the calculations, some of which are the result of a considerable amount of work. (1) No shortening of the work is possible by recourse to complex notation, and the chance of a mistake is thereby considerably increased. (2) The process of solution is rendered more difficult if the magnetic field is taken as zero. The result for zero magnetic field may always be obtained by letting ω tend to zero in the final solution. (3) In the case $U_0=0$, $X_0=0$ alternative methods of arriving at the general solution may suggest themselves. The method given in the text for the solution of equation (98) has the advantage of being strictly orthodox.

---

* Megaw, *loc. cit.*

§ 5. GENERAL SOLUTION

(5.1) *Reduction of primary equation to linear form.* No method has yet been found for solving directly for U_x, equation (43), except (4.1) in the steady state. It is, of course, possible to find variables τ, t_0 such that

$$\frac{\partial}{\partial t}\bigg|_x + U_x\frac{\partial}{\partial x}\bigg|_t = \frac{\partial}{\partial \tau}\bigg|_{t_0}.$$

But in this case τ is the instantaneous transit time and the solution obtained will be the same as that for individual electrons, in which the arguments of the circular functions all have a fluctuating component at a given value of x. Now we are particularly anxious to avoid this feature, and we therefore seek to express equation (43) in terms of variables of which one shall be independent of time at a given value of x. Now equation (43) as it stands does satisfy this condition, since one of the independent variables is x itself. That we seek a change of variables from x and t is due entirely to the fact that the operator $(U_x\,\partial/\partial x)$ does not apply when t is constant, since U_x is itself a function of t; alternatively if we regard $\dfrac{\partial}{\partial x}\bigg|_t$ as the operator, we are faced with a partial differential equation of the third degree in U_x. τ

The only method available for achieving the desired results is that of harmonic subdivision already described in (2.1) and (2.3). We shall then merely have to find variables T, t_0 such that

$$\frac{\partial}{\partial t}\bigg|_x + \bar{U}\frac{\partial}{\partial x}\bigg|_t = \frac{\partial}{\partial T}\bigg|_{t_0} \qquad \text{......(93)},$$

a relatively simple matter, since $\bar{U}$ is independent of t and T will therefore be the steady state value of the transit time to the plane x. We can in fact replace $\bar{U}\,\partial/\partial x$ by $\partial/\partial T$ so that equation (93) becomes

$$\frac{\partial}{\partial t}\bigg|_T + \frac{\partial}{\partial T}\bigg|_t = \frac{\partial}{\partial T}\bigg|_{t_0} \qquad \text{......(94)}.$$

The following analysis suffices to show that t_0 represents the instant at which electrons leave the cathode. Let f be any function of t, T and let $\bar{f}$ be the same function when expressed in terms of t_0, T'. $f, \bar{f}$

Then $$\frac{\partial f}{\partial t}\delta t + \frac{\partial f}{\partial T}\delta T = \frac{\partial \bar{f}}{\partial t_0}\delta t_0 + \frac{\partial \bar{f}}{\partial T'}\delta T'.$$

Writing $$\delta T = \frac{\partial t}{\partial t_0}\delta t_0 + \frac{\partial t}{\partial T'}\delta T',$$

we find that $$\frac{\partial f}{\partial t}\left(\frac{\partial t}{\partial t_0}\delta t_0 + \frac{\partial t}{\partial T'}\delta T'\right) + \frac{\partial f}{\partial T}\delta T = \frac{\partial \bar{f}}{\partial t_0}\delta t_0 + \frac{\partial \bar{f}}{\partial T'}\delta T'.$$

Since $\delta T' = \delta T$ we have, comparing terms in δT,

$$\frac{\partial f}{\partial t}\frac{\partial t}{\partial T'} + \frac{\partial f}{\partial T} = \frac{\partial \bar{f}}{\partial T'}.$$

The function f is seen to satisfy equation (94) provided

$$\frac{\partial t}{\partial T'}=1,$$

which gives on integrating partially with respect to T, t_0 being constant

$$t=T'+[t_0],$$

or writing $T'=T$ we obtain $\qquad t_0=t-T.$

Since $T=0$ at the cathode, t_0 represents the instant at which the electrons left the cathode. Thus one of our independent variables is now the same as in (2.4), namely t_0. The other independent variable T differs from the $(t-t_0)$ of (2.4) in that the latter represents instantaneous transit time: at a given value of x the $(t-t_0)$ of (2.4) contains a fluctuating component, whereas T is constant. We may thus think of T as x itself, and no confusion will arise if we refer to the plane x as *the plane* T, bearing in mind the slight spread in T_1, see (4.3).

We shall require to replace U_x by $\overline{U}+u+\dots$ in equation (43), which gives

$$\left[\frac{\partial}{\partial t}+(\overline{U}+u+\dots)\right]^2(\overline{U}+u+\dots)=4\pi\frac{e}{m}c^2J-\omega^2(\overline{U}+u+\dots).$$

Applying equations (93) and (94) we obtain

$$\left[\left(1+\frac{u}{\overline{U}}\right)\frac{\partial}{\partial T}\bigg|_{t_0'}-\frac{u}{\overline{U}}\frac{\partial}{\partial t}\bigg|_{T}\right]^2(\overline{U}+u)=4\pi\frac{e}{m}c^2J-\omega^2(\overline{U}+u) \quad\dots\dots(95).$$

ψ_1 Let us extract all terms containing u once only. Let the sum of these be ψ_1. Then, since $\partial\overline{U}/\partial t=0$, we have

$$\psi_1=\frac{\partial^2 u}{\partial T^2}\bigg|_{t_0'}+\left(\frac{1}{\overline{U}}\frac{d\overline{U}}{dT}\right)\frac{\partial u}{\partial T}\bigg|_{t_0'}+\left\{\frac{2}{\overline{U}}\frac{d^2\overline{U}}{dT^2}-\frac{1}{\overline{U}^2}\left(\frac{d\overline{U}}{dT}\right)^2+\omega^2\right\}u \quad\dots\dots(96).$$

ψ_2 We now assume a known value for J of the type $J=\bar{J}+\sum_1^r j_n\sin^n pt$. If ψ_2 denote
ψ_3 the sum of all terms containing u twice and over ψ_3 the sum of all terms containing u three times over, equation (93) may be resolved into discrete equations of zero, first, second, third, ... order respectively

$$\frac{d^2\overline{U}}{dT^2}+\omega^2\overline{U}=4\pi\frac{e}{m}c^2\bar{J} \quad\dots\dots(44a),$$

$$\psi_1=4\pi\frac{e}{m}c^2j_1\sin p\,(t_0'+T) \quad\dots\dots(97),$$

$$\psi_1'+\psi_2=4\pi\frac{e}{m}c^2j_2\sin^2 p\,(t_0'+T),$$

$$\psi_1''+\psi_{12}'+\psi_3=4\pi\frac{e}{m}c^2j_3\sin^3 p\,(t_0'+T),$$

$$\left[\begin{array}{l}\psi_1'''+\psi_{12}''+\psi_{22}'+0=4\pi\dfrac{e}{m}c^2j_4\sin^4 pt\\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\end{array}\right],$$

$_1', \psi_1'', u', u''$ ψ_1', ψ_1'' correspond to the second and third-order velocities u', u'' which, for the sake of clearness, were omitted from equation (95). ψ_n' represents the sum of third-order terms containing u once and u' once. There is no difficulty in obtaining ψ_2, ψ_2' and ψ_3 but, as was pointed out in (2.3), we shall only require the fundamental. ψ_n'

Writing $\omega T=\theta$ and giving $\overline{U}$ its value from equation (51), we obtain from equations (96) and (97) the equation

$$\frac{\partial^2 u}{\partial \theta^2}+\sigma_1\frac{\partial u}{\partial \theta}+\frac{1-n_1^2-n_2^2}{(1-n_1\cos\theta+n_2\sin\theta)^2}u=\frac{\beta}{\omega^2}\sin\left(pt_0+\frac{\theta}{\omega}\right) \qquad \text{......(98)},$$

in which

$$\left.\begin{aligned}\frac{\partial}{\partial\theta}&\equiv\left.\frac{\partial}{\partial\theta}\right|_{t_0}\\ \sigma_1&\equiv\frac{1}{\overline{U}}\frac{d\overline{U}}{d\theta}=\frac{n_1\sin\theta+n_2\cos\theta}{1-n_1\cos\theta+n_2\sin\theta}\\ \beta&\equiv 4\pi\frac{e}{m}c^2 j_1\end{aligned}\right\} \qquad \text{......(99)}.$$

σ_1

β

(5.2) *Derivation of general solution.* In attempting to solve equation (98) we are assisted by our knowledge of simpler cases, which prompts us to notice that σ_1 is itself a solution to the equation when $\beta=0$. When any one solution has been obtained for the equation for $\beta=0$, it is a simple matter to obtain the general solution for $\beta=0$. This is found to be

$$\left.\begin{aligned}&A\sigma_2+B\sigma_1\\ \text{where}\qquad &\sigma_2\equiv\frac{n_1^2+n_2^2-n_1\cos\theta+n_2\sin\theta}{1-n_1\cos\theta+n_2\sin\theta}\end{aligned}\right\} \qquad \text{......(100)}.$$

σ_2

The general solution in case $\beta\neq 0$ then proceeds as follows. Let us change the variable to η where

$$u=\sigma_1\eta \qquad \text{......(101)}.$$

Equation (98) becomes, after being divided through by σ_1,

$$\frac{\partial^2\eta}{\partial\theta^2}+\left(\sigma_1+\frac{2}{\sigma_1}\frac{d\sigma_1}{d\theta}\right)\frac{\partial\eta}{\partial\theta}=\frac{\lambda}{\sigma_1}\sin(pt_0+r\theta) \qquad \text{......(102)},$$

where

$$r\equiv\frac{p}{\omega},\quad \lambda\equiv\frac{\beta}{\omega^2} \qquad \text{......(103)}.$$

r, λ

Since equation (102) contains no term in η, the integrating factor is

$$e^{\int\left(\sigma_1+\frac{2}{\sigma_1}\frac{d\sigma_1}{d\theta}\right)d\theta} \text{ which } =\sigma_1^2\, e^{\int\sigma_1 d\theta}=\sigma_1^2(1-n_1\cos\theta+n_2\sin\theta).$$

Solving for $\partial\eta/\partial\theta$ we obtain, writing $A(n_1^2+n_2^2)$ for the constant of integration, A

$$\frac{\partial\eta}{\partial\theta}=\frac{1-n_1\cos\theta+n_2\sin\theta}{(n_1\sin\theta+n_2\cos\theta)^2}\left[A(n_1^2+n_2^2)+\lambda\int(n_1\sin\theta+n_2\cos\theta)\sin(pt_0+r\theta)\,d\theta\right] \qquad \text{......(104)}.$$

The integral appearing in equation (104) presents no difficulty, but the next stage will involve rather unusual integrals. These have, however, given little trouble and can all be evaluated in circular functions. For example

$$\int\frac{1-n_1\cos\theta+n_2\sin\theta}{(n_1\sin\theta+n_2\cos\theta)^2}\,d\theta=\frac{N^2-n_1\cos\theta+n_2\sin\theta}{N^2(n_1\sin\theta+n_2\cos\theta)}.$$

σ_3 Writing σ_3 for the quantity $\dfrac{N^2-n_1\cos\theta+n_2\sin\theta}{n_1\sin\theta+n_2\cos\theta}$, we also require

$$\frac{N^2}{r-1}\int\frac{1-n_1\cos\theta+n_2\sin\theta}{(n_1\sin\theta+n_2\cos\theta)^2}\sin\,[pt_0+(r-1)\,\theta]\,d\theta$$
$$=\frac{\sigma_3\sin\,(pt-\theta)}{r-1}-\int\sigma_3\cos\,[pt_0+(r-1)\,\theta]\,d\theta$$

and

$$\int\sigma_3\,(n_1\sin\theta+n_2\cos\theta)\sin\,(pt_0+r\theta)\,d\theta=\int(N^2-n_1\cos\theta+n_2\sin\theta)\sin\,(pt_0+r\theta)\,d\theta,$$

N where $$N^2\equiv n_1{}^2+n_2{}^2,\qquad \sigma_3=\frac{\sigma_2}{\sigma_1};\qquad t=t_0+\frac{r\theta}{p}\qquad\text{......(105).}$$

With the help of the above integrals we arrive at the equation

$$\eta=A\sigma_3+B+\frac{\lambda}{r}\cos pt+\frac{\lambda}{2N^2}\left[(n_1\sigma_3-n_2)\left\{\frac{\sin\,(pt-\theta)}{r-1}-\frac{\sin\,(pt+\theta)}{r+1}\right\}\right.$$
$$\left.-(n_2\sigma_3+n_1)\left\{\frac{\cos\,(pt-\theta)}{r-1}+\frac{\cos\,(pt+\theta)}{r+1}\right\}\right],$$

which simplifies to $$\eta=A\sigma_3+B-\frac{\lambda}{r^2-1}\left[\frac{\sin pt}{\sigma_1}+\frac{\cos pt}{r}\right]\qquad\text{......(106).}$$

To obtain u we multiply by σ_1, equation (101). Giving λ and r their values from equation (103), we obtain the following general solution for u:

$$u=A\sigma_2+B\sigma_1-\frac{\beta}{p^2-\omega^2}\left[\sin pt+\frac{\omega}{p}\sigma_1\cos pt\right]\qquad\text{......(107),}$$

in which A and B are arbitrary functions of t_0 and σ_1, σ_2 are given by equations (99) and (100) respectively.

(5.3) *Notes on the general solution.* It is immaterial whether we regard u as a function of θ and t_0 or of θ and t or of all three variables θ, t_0 and t. We shall in any case retain θ, inasmuch as σ_1 and σ_2 are expressed in terms of θ. Comparison with equation (14) reveals that the factor $(p^2-\omega^2)$ is common to both solutions. In other respects there is little obvious resemblance between equations (14) and (107). The solutions are nevertheless equivalent, the large apparent difference arising from the different meaning attached to t in the two coordinate systems, as has been pointed out in (5.1).

We note that the form of equation (107) is similar to that obtained in the very much simpler case of zero initial velocities, accelerations, and magnetic field*. To obtain this case all we need to do is to write

$$\left.\begin{aligned}&\sigma_2=1,\qquad \sigma_1=\frac{\sin\theta}{1-\cos\theta},\\ &\text{so that}\qquad \underset{\theta\to0}{\mathrm{Lt}}\left(\frac{\omega}{p}\sigma_1\right)=\frac{2\omega}{p\theta}=\frac{2}{pT}=\frac{2}{\xi}\end{aligned}\right\}\qquad\text{......(108).}$$

---

* *Phil. Mag.* **5** (1928), equation (19), p. 649.

§6. BOUNDARY CONDITIONS

(6.1) *General.* It is usual at this stage to declare that the mathematics of the problem is at an end, for the determination of arbitrary constants and of arbitrary functions is governed by purely physical considerations. The determination of A and B in equation (107) is, however, not devoid of mathematical interest. Inspection of equation (107) reveals the existence of singularities when $p=\pm\omega$ and also when $p=0$. It will be seen that these singularities can be removed by choosing the arbitrary functions in such a way that infinities do not occur in u either when $p=\pm\omega$ or when $p=0$.

Any suggestion that an infinity might occur when $p=\omega$ is refuted by working out the solution with p equal to ω at the outset, when it is found that no sign of resonance occurs. Experiment, moreover, confirms this point. In general nothing unusual occurs experimentally either when $p=\omega$ or when $p=0$, unless θ has certain values. We therefore seek to remove the singularities for all values of θ, including those which are favourable to electronic oscillations. For even in this case the alternating component of velocity cannot be infinite though, as we shall see, it may assume large values.

The number of physical conditions we shall need will not exceed the number of arbitrary constants, provided such conditions are truly independent. Now it is an experimental fact that reversal of the current in the field coil of the electromagnet is without effect on the behaviour of a magnetron, apart from experimental errors such as slight lack of alignment, and we may thus introduce as a condition governing A and B the assertion that the solution for u remains unchanged when ω is changed to $-\omega$. We cannot then, however, regard the condition $p=\pm\omega$ as providing two independent physical conditions. The condition $p=0$ is an independent physical condition.

We have thus far three physical conditions wherewith to determine A and B. But the most general form of A and B compatible with the problem involves four arbitrary constants:

$$\left.\begin{aligned} A&=A_1\sin pt_0+A_2\cos pt_0\\ B&=B_1\sin pt_0+B_2\cos pt_0\end{aligned}\right\} \quad\ldots\ldots(109).$$

We shall accordingly require a fourth condition in order to determine A and B completely, unless it should so happen that A_1, A_2, B_1, B_2 are in some way related, so that they cannot all be regarded as arbitrary. While it will eventually appear that simple relations exist between the A's and the B's, there is not sufficient evidence at the present stage to assume that fewer than four arbitrary constants exist. We therefore have to seek a further physical condition which must also be independent of the three already obtained. Now while the solution must certainly remain finite when $\omega=0$, this condition is already satisfied in equation (107). Suppose, however, it should happen that in the satisfaction of other conditions terms were introduced with singularities at $\omega=0$; we should be justified in adjusting such terms in accordance with the condition $\omega=0$. It will appear that this state of affairs does arise, and it does seem probable that the condition $\omega=0$ is of paramount importance.

Now the acceleration of electrons must never be infinite, but in regard to the avoidance of infinities, care must be exercised in ensuring that any conditions imposed shall be truly independent. The question we have now to consider is whether the condition that the velocity shall never be infinite automatically excludes the possibility of an infinite acceleration. This brings us back to the old enigma of the irresistible force acting on the immovable body. Now whatever happens in this case, the irresistible force could be achieved by causing a finite momentum to suffer a change in an infinitesimal time. If, moreover, the change in momentum be brought about by a change of density rather than by a change of velocity, the velocity must remain finite, since by hypothesis it was finite before the change took place.

More than sufficient has been said to establish the fact that an exceedingly high value of acceleration could occur with an exceedingly low value of velocity, and we therefore see that by applying our conditions to the electron-acceleration the number of available physical conditions is doubled.

When the electrons leave a hot cathode they do so with a distribution of velocities, but whatever the velocity of an electron at the instant of emergence the alternating component of velocity must then be zero. Now the strange thing about this condition is the fact that it has been found of no assistance whatever in the determination of the arbitrary functions in the general case now under consideration. All attempts to arrive at the final solution on the basis that u must vanish at the hot cathode proved unavailing; yet, when the final solution was obtained on the basis of conditions connected with the avoidance of infinities in u, the condition $u=0$ at the cathode was found on inspection to be satisfied. The explanation of this apparent anomaly is as follows. The condition $u=0$ at the cathode is actually inherent in the analysis owing to the non-linearity of the primary equation, together with the introduction at an early stage of the assumed boundary values of d.-c. velocity and acceleration and of the total current. Llewellyn in a recent paper* has proposed a general form of boundary condition for u in the simple case where emission velocities are assumed to be zero. One object of Llewellyn's extra condition imposed on u was to enable the analysis to be extended to cover the case of virtual cathodes, where there is strong experimental evidence that the alternating velocity and acceleration must be very different from zero. It appears, however, from the present paper that the value of u is completely determined at a virtual cathode, no less than at a real one, without the imposition of any conditions on u, apart from those connected with the avoidance of singularities. Llewellyn's paper should be consulted for the opposite point of view, which also appears at first sight to be the natural and correct one to adopt.

Summarizing the above discussion we may say that any boundary conditions imposed on u, other than the following, are irrelevant: (i) when $p=\pm\omega$ the solution remains finite at all values of T; (ii) when ω is changed to $-\omega$ the solution remains unchanged; (iii) when $p=0$ the solution remains finite at all values of T; (iv) when $\omega=0$ the solution remains finite at all values of T.

---

* *Proc. Inst. Radio Engrs*, N.Y., **21**, 1532 (1933).

We also have corresponding conditions imposed on the acceleration, but it appears that A and B are completely determined by conditions (i) to (iv) above.

(6.2) *Determination of arbitrary functions.* Inspection of the general solution, e.g.

$$u = A\sigma_2 + B\sigma_1 - \frac{\beta}{p^2-\omega^2}\sin pt - \frac{\beta}{p^2-\omega^2}\frac{\omega}{p}\sigma_1\cos pt \qquad \ldots\ldots(107a),$$

reveals that the last term has singularities when $p=0$ and also when $p=\omega$. It is therefore desirable to make use of the identity

$$\frac{\omega}{p} = \frac{p}{\omega} - \frac{p^2-\omega^2}{\omega p} \qquad \ldots\ldots(110),$$

so that we obtain

$$u = A\sigma_2 + B\sigma_1 - \frac{\beta}{p^2-\omega^2}\sin pt - \frac{\beta}{p^2-\omega^2}\frac{p}{\omega}\sigma_1\cos pt + \frac{\beta\sigma_1}{\omega p}\cos pt \qquad \ldots\ldots(107b).$$

We have now separated out the singularity occurring when $p=0$, so that the one term which stands in need of condition (iii) does not now require condition (i). It is true that in effecting such a separation we have produced two terms with singularities when $\omega=0$, whereas none existed before. We know, however, that these two terms balance one another, so that all we have to remember is that any values assigned to A and B must not lead to a singularity in $A\sigma_2 + B\sigma_1$ when $\omega=0$.

Condition (iii): Concerning ourselves with the removal of the singularity which occurs in the last term when $p=0$, we find that the necessary and sufficient condition that u shall remain finite when $p=0$ is that the arbitrary function B shall contribute

$$-\frac{\beta}{\omega p}\cos pt_0{}^{*} \qquad \ldots\ldots(111).$$

Condition (i): Concerning ourselves next with the removal of the singularities in the two terms of equation (107b) having $p^2-\omega^2$ in the denominator, we have to give σ_1 and σ_2 their values from equations (99) and (100) and select the coefficient of $\sin pt_0$ and also the coefficient of $\cos pt_0$. Since for u to remain finite when $p=\pm\omega$, $u(p^2-\omega^2)$ must vanish, the above-mentioned coefficients are respectively equated to zero for the case $p=\pm\omega$.

Since factors of the type $(p/\pm\omega)^q$ may be introduced with impunity without affecting condition (i) *per se*, we have to decide the value of q with the assistance of conditions (ii), (iii) and (iv). We thus find that A and B contribute as follows:

$$\left.\begin{aligned} A\text{'s contribution} &= \frac{\beta}{p^2-\omega^2}\frac{1}{N^2}\left(-n_1\sin pt_0 + n_2\frac{p}{\omega}\cos pt_0\right) \\ B\text{'s contribution} &= \frac{\beta}{p^2-\omega^2}\frac{1}{N^2}\left(n_2\sin pt_0 + n_1\frac{p}{\omega}\cos pt_0\right) \end{aligned}\right\} \qquad \ldots\ldots(112).$$

The quantities σ_1 and n_2 change sign when ω changes sign, while σ_2 and n_1 remain unchanged. Let us now confirm that condition (ii) is satisfied, i.e. that the solution remains unaffected when ω is changed to $-\omega$.

---

[*] It must be noted that whereas T is always a finite interval, t_0 and t may approach infinity since the choice of time-origin is arbitrary. In taking limits when $p\to 0$ we must therefore always leave $\sin pt$ and $\cos pt$ intact.

We have, if equations (111) and (112) represent the entire contributions of A and B:

$$\begin{aligned}u=\frac{\beta}{p^2-\omega^2}\frac{1}{N^2}&\left(-n_1\sin pt_0+n_2\frac{p}{\omega}\cos pt_0\right)\sigma_2\\&+\frac{\beta.}{p^2-\omega^2}\frac{1}{N^2}\left(n_2\sin pt_0+n_1\frac{p}{\omega}\cos pt_0\right)\sigma_1\\&-\frac{\beta}{p^2-\omega^2}\left(\sin pt+\frac{p}{\omega}\sigma_1\cos pt\right)+\frac{\beta}{\omega p}\sigma_1(\cos pt-\cos pt_0)\quad\ldots\ldots(113).\end{aligned}$$

Changing ω to $-\omega$, n_2 to $-n_2$ and σ_1 to $-\sigma_1$, we obtain

$$\begin{aligned}u=\frac{\beta}{p^2-\omega^2}\frac{1}{N^2}&\left[\left(-n_1\sin pt_0-n_2\frac{p}{-\omega}\cos pt_0\right)\sigma_2-\left(n_2\sin pt_0+n_1\frac{p}{\omega}\cos pt_0\right)(-\sigma_1)\right]\\&-\frac{\beta}{p^2-\omega^2}\left(\sin pt+\frac{p}{-\omega}(-\sigma_1)\cos pt\right)+\frac{\beta(-\sigma_1)}{-\omega p}(\cos pt-\cos pt_0),\end{aligned}$$

showing that u remains unaltered.

We now confirm that condition (iv) is satisfied in equation (113). There are four terms in equation (112) with singularities when $\omega=0$.

We have

$$\operatorname*{Lt}_{\omega\to0}(n_2)=0,$$

$$\operatorname*{Lt}_{\omega\to0}(n_1)=1=\operatorname*{Lt}_{\omega\to0}(N^2),$$

$$\left.\begin{aligned}&\text{whence}\quad\operatorname*{Lt}_{\omega\to0}\left[\frac{\beta\sigma_1}{p^2-\omega^2}\frac{n_1}{N^2}\frac{p}{\omega}\cos pt_0-\frac{\beta}{\omega p}\sigma_1\cos pt_0\right]=0\\&\text{and}\quad\operatorname*{Lt}_{\omega\to0}\left[\frac{-\beta\sigma_1}{p^2-\omega^2}\frac{p}{\omega}\cos pt+\frac{\beta}{\omega p}\sigma_1\cos pt\right]=0\end{aligned}\right\}\quad\ldots\ldots(114).$$

Thus condition (iv) is satisfied. But since in arriving at equations (111) and (112) all four conditions have been used, A and B must be completely determined, and equation (113) must be the final solution for u. In view of equation (110), equation (113) may now be written

$$\begin{aligned}u=\frac{\beta}{p^2-\omega^2}\Bigg[&-\sin pt-\frac{n_1\sigma_2-n_2\sigma_1}{N^2}\sin pt_0+\frac{\omega}{p}\sigma_1(\cos pt_0-\cos pt)+\frac{p}{\omega}\frac{n_2\sigma_2}{N^2}\cos pt_0\\&-\frac{(N^2-n_1)}{N^2}\frac{p}{\omega}\sigma_1\cos pt_0\Bigg]\quad\ldots\ldots(113a).\end{aligned}$$

§7. DISCUSSION OF FINAL SOLUTION

(7.1) *General.* For purposes of comparison with simpler cases it is convenient to separate out terms correcting for initial velocities and accelerations. We then obtain

$$\begin{aligned}u=&\frac{\beta}{p^2-\omega^2}\left[-\sin pt-\sigma_2\sin pt_0+\frac{\omega}{p}\sigma_1(\cos pt_0-\cos pt)\right]\\&+\frac{\beta}{p^2-\omega^2}\left\{\frac{(N^2-n_1)\sigma_2+n_2\sigma_1}{N^2}\sin pt_0-\frac{p}{\omega}\frac{(N^2-n_1)\sigma_1-n_2\sigma_2}{N^2}\cos pt_0\right\}\quad\ldots\ldots(113b).\end{aligned}$$

In the event of initial velocities and accelerations being zero the terms between curly brackets reduce to zero, while the terms between square brackets remain. Except for very small values of θ we may write

$$\sigma_1 \doteqdot \frac{\sin\theta}{1-\cos\theta}; \quad \sigma_2 \doteqdot 1.$$

We thus obtain for zero initial velocities and accelerations

$$u=\frac{\beta}{p^2-\omega^2}\left[-\sin pt-\sin pt_0+\frac{\omega}{p}\frac{\sin\theta}{1-\cos\theta}(\cos pt_0-\cos pt)\right] \quad \text{......(113}c\text{)},$$

which is seen to correspond with equation (20) of my 1928 paper, but with $\frac{2}{\xi}$ replaced by $\frac{\omega}{p}\frac{\sin\theta}{1-\cos\theta}$. Actually we shall see immediately that $\sigma_2 < 1$ at points very near the cathode, so that for this reason alone equation (113c) cannot be applied near the cathode.

As a searching test on the uniqueness of equation (113b) let us seek the value of u at the cathode. We have to let T tend to zero, so that

$$\underset{T\to 0}{\text{Lt}}\,(\sigma_2)=\frac{N^2-n_1}{1-n_1}=-n_1+\frac{n_2{}^2}{1-n_1} \quad \text{......(115)},$$

$$\underset{T\to 0}{\text{Lt}}\,(\sigma_1)=\frac{n_2}{1-n_1} \quad \text{......(116)}.$$

When t_0 is equated to t equation (113b) becomes

$$\begin{aligned} u&=\frac{\beta}{p^2-\omega^2}\left[-\sin pt-\frac{N^2-n_1}{1-n_1}\sin pt\right] \\ &\quad +\frac{\beta}{p^2-\omega^2}\left\{\frac{(N^2-n_1)^2+n_2{}^2}{N^2(1-n_1)}\sin pt-\frac{p}{\omega}\frac{(N^2-n_1)\,n_2-(N^2-n_1)\,n_2}{N^2(1-n_1)}\cos pt\right\} \\ &=\frac{\beta}{p^2-\omega^2}\left[\frac{-N^2+N^2n_1-N^4+n_1N^2+N^4-2N^2n_1+n_1{}^2+n_2{}^2}{N^2(1-n_1)}\right]\sin pt+[0]\cos pt \\ &=0. \end{aligned}$$

We thus see that the inclusion of initial velocities and accelerations provides all the boundary conditions required at the cathode: the value of u, the alternating velocity, reduces to zero at the cathode without the imposition of any such boundary condition on u. All attempts to arrive at equation (113) with $u=0$ as one of the four boundary conditions met with no success.

From consideration of equation (115) it is seen that the value of σ_2 at the cathode is approximately -1. It is exactly -1 at the potential-minimum. This is an interesting state of affairs which corresponds to a change of sign in σ_2, where

$$\sigma_2=\frac{N^2-n_1\cos\theta+n_2\sin\theta}{1-n_1\cos\theta+n_2\sin\theta}$$

at some point which must lie close to the cathode. Under these circumstances, giving n_1, n_2 and N^2 their values from equations (50) and (105) and writing $\theta=\omega T$, we obtain when ωT is small

$$\sigma_2 = \frac{\left(\frac{eX_0}{mC}\right)^2 - \frac{U_0}{C} + \frac{eX_0}{mC}T + \frac{T^2}{2}}{\frac{U_0}{C} + \frac{eX_0}{mC}T + \frac{T^2}{2}}.$$

T_m Writing T_m for the time taken by electrons to travel from cathode to potential-minimum, we have in accordance with equation (83):

$$-\frac{eX_0}{mC} = T_m,$$

so that

$$\sigma_2 = \frac{T_m{}^2 - U_0/C - T_m T + \tfrac{1}{2}T^2}{U_0/C - T_m T + T^2/2} \qquad \text{......(117)},$$

which has its minimum value (-1) when $T = T_m$, i.e. at the potential-minimum. If, however, $\sigma_2 = 0$ we obtain

$$T = T_m + \sqrt{\left(\frac{2U_0}{C} - T_m{}^2\right)},$$

so that it is not at the potential-minimum that σ_2 changes sign but at some point further removed from the cathode, probably nearly at $x = 2x_m$. If the potential-minimum coincides with the cathode, then the approximate solution equation (113) applies at the cathode with $-\sin pt_0$ replaced by $+\sin pt_0$, thus removing any doubt that u may not vanish at the cathode according to equation (113*c*). Actually it may be regarded as somewhat remarkable that equation (113*c*) as it stands does vanish at the cathode, for if we place $t = t_0$ straight away we obtain

$$u_0 = -\frac{\beta}{p^2 - \omega^2}\, 2 \sin pt_0 \qquad \text{......(118)}.$$

On taking limits we find, however, a contribution from the terms in $\cos pt$, $\cos pt_0$ which exactly balance equation (118) if σ_1 is taken as $2/\theta$ at the cathode. This, in turn, is now seen to be impermissible since, writing σ_1 in full, we have when ω is small

$$(\omega\sigma_1) = \frac{\frac{e}{m}X_0 + CT}{U_0 + \frac{e}{m}X_0\, T + \tfrac{1}{2}CT^2} \qquad \text{......(119)}.$$

Only if we write $X_0 = 0 = U_0$ *before* equating T to zero can we obtain $\sigma_1 = 2/\omega T$, and clearly the true value of σ_1 at the cathode is

$$(\sigma_1)_0 = \frac{eX_0}{m\omega U_0}.$$

For the case of zero magnetic field, for which σ_1 becomes infinite, the interesting quantities in the solution for u are σ_2 and $\underset{\omega\to 0}{\mathrm{Lt}}\,(\omega\sigma_1)$. If we adopt the typical values of

C, T_m, X_0 and U_0 worked out in (4.6), equation (87), and table 1, e.g.

$$\begin{aligned} C &= 0{\cdot}9987 \times 10^{27}\ \text{cm./sec}^3, \\ T_m &= 2{\cdot}0154 \times 10^{-10}\ \text{sec.}, \\ X_0 &= 113{\cdot}96 \times 10^{8}\ \text{e.m.u.}, \\ U_0 &= 2{\cdot}2330 \times 10^{7}\ \text{cm./sec.}, \end{aligned}$$

we obtain, using equations (117) and (119),

$$\sigma_2 = \frac{(1{\cdot}8254 \times 10^{-20}) - (2{\cdot}0154 \times 10^{-10}\, T) + 0{\cdot}5 T^2}{(2{\cdot}2360 \times 10^{-20}) - (2{\cdot}0154 \times 10^{-10}\, T) + 0{\cdot}5 T^2} \qquad \ldots\ldots(117a),$$

$$\operatorname*{Lt}_{\omega \to 0} (\omega\sigma_1) = \frac{-2{\cdot}0127 \times 10^{17} + 0{\cdot}9987 \times 10^{27}\, T}{(2{\cdot}2330 \times 10^{7}) - (2{\cdot}0127 \times 10^{17} T) + (0{\cdot}4999 \times 10^{27} T^2)} \qquad \ldots\ldots(119a).$$

If we give T its value corresponding to transit time between cathode and anode, we have, taking $V = 300$ volts and $d = 0{\cdot}5$ cm. from table 1:

$$T = \frac{3d}{\sqrt{(-2Ve/m)}} = \frac{1{\cdot}5}{\sqrt{(3{\cdot}54 \times 3 \times 10^{17})}} = 1{\cdot}4157 \times 10^{-9}\ \text{sec.}$$

We obtain further

$$\sigma_2 = \frac{1{\cdot}8254 \times 10^{-20} - 2{\cdot}8529 \times 10^{-19} + 1{\cdot}0020 \times 10^{-18}}{2{\cdot}2360 \times 10^{-20} - 2{\cdot}8529 \times 10^{-19} + 1{\cdot}0020 \times 10^{-18}} = 0{\cdot}9945,$$

$$\operatorname*{Lt}_{\omega \to 0} (\omega\sigma_1) = \frac{-2{\cdot}0127 \times 10^{17} + 1{\cdot}4138 \times 10^{18}}{2{\cdot}2330 \times 10^{7} - 2{\cdot}8490 \times 10^{8} + 1{\cdot}0006 \times 10^{9}} = 1{\cdot}64 \times 10^{9}.$$

The above values are to be compared with the corresponding values obtained when a magnetic field of critical value is applied: under these conditions it is readily shown that at the anode

$$\sigma_2 = -1 + \frac{\omega^2 U_0}{C} + \frac{\left(\frac{e}{m} X_0\right)^2}{U_0 C} = 0{\cdot}940,$$

$$\omega\sigma_1 = \frac{eX_0}{mU_0} = -9{\cdot}02 \times 10^{9},$$

which shows that the value of σ_2 at the anode is not greatly affected by a magnetic field, but that $\omega\sigma_1$ has a negative instead of a positive value. By comparison with equation (119a) we see that the value of $\omega\sigma_1$ at the anode in the critical condition is the same as its value at the cathode. But from equation (99)

$$\omega\sigma_1 = \frac{\omega}{U}\frac{dU}{d\theta} = \frac{1}{U}\frac{dU}{dT}.$$

In the critical condition U has the same value at the anode as at the cathode, namely U_0. Thus dU/dT also must have the same value in both places, namely $(e/m)\,X_0$. We see therefore that a virtual cathode exists just in front of the anode having just the same properties as the real cathode in respect of the forward component of direct-current velocity.

u_c (7.2) *Alternating velocity at critical plane.* When we come to consider the alternating velocity, we find, writing u_c for the value of u at the anode in the critical condition,

$$u_c=\frac{\beta}{p^2-\omega^2}\left[\left(1+\cos\frac{2\pi p}{\omega}+\frac{n_2}{1-n_1}\frac{\omega}{p}\sin\frac{2\pi p}{\omega}\right)\sin pt\right.$$
$$\left.-\left\{\sin\frac{2\pi p}{\omega}+\frac{n_2}{1-n_1}\frac{\omega}{p}\left(1-\cos\frac{2\pi p}{\omega}\right)\right\}\cos pt\right]\text{......(120)},$$

in which

$$\frac{n_2}{1-n_1}=\frac{eX_0}{m\omega U_0}=-\frac{X_0}{U_0H_c},$$

where H_c and X_0 are both positive.

Table 3 contains useful particular cases of the above expression corresponding to different values of p/ω.

Table 3

p/ω	u_c
0	$\frac{+\beta}{\omega^2}\left(\frac{2\pi X_0}{U_0H_c}\right)\sin pt$
$\frac{1}{4}$	$\frac{16}{15}\frac{\beta}{\omega^2}\left[\left(1+\frac{4X_0}{U_0H_c}\right)\sin pt+\left(1-\frac{4X_0}{U_0H_c}\right)\cos pt\right]$
$\frac{1}{2}$	$\frac{8}{3}\frac{\beta}{\omega^2}\left[\sin pt-\frac{2X_0}{U_0H_c}\cos pt\right]$
$\frac{3}{4}$	$\frac{16}{7}\frac{\beta}{\omega^2}\left[\left(1-\frac{4X_0}{3U_0H_c}\right)\sin pt-\left(1+\frac{4X_0}{3U_0H_c}\right)\cos pt\right]$
1	$\frac{-\pi\beta}{\omega^2}\left[\cos pt+\frac{X_0}{U_0H_c}\sin pt\right]$
$\frac{5}{4}$	$\frac{16}{9}\frac{\beta}{\omega^2}\left[\left(\frac{-4X_0}{5U_0H_c}-1\right)\sin pt+\left(\frac{4X_0}{5U_0H_c}-1\right)\cos pt\right]$
$1\frac{1}{2}$	$\frac{8}{5}\frac{\beta}{\omega^2}\left[\frac{2X_0}{3U_0H_c}\cos pt-\sin pt\right]$
$\frac{7}{4}$	$\frac{16}{33}\frac{\beta}{\omega^2}\left[\left(1+\frac{4X_0}{7U_0H_c}\right)\cos pt-\left(1-\frac{4X_0}{7U_0H_c}\right)\sin pt\right]$
2	0

The alternating velocity u_c as given by equation (120) should strictly be regarded in the first instance as corresponding to the critical plane rather than to the anode. The critical condition corresponds, however, to the electrons coming to rest exactly at the anode, so that the distinction is apparent rather than real. The point was raised in order to prepare the way for a further conclusion which requires a little consideration. The question to be settled is as follows. Does the existence of an alternating component of velocity superimposed upon the steady component of velocity in any way alter the conditions under which equation (120) was derived?

In the first place we noted that whatever the value of u_c, the value of $\overline{U}$ remains unaffected. This may be shown by referring back to equations (44*a*) and (97).

Thus the value of θ is 2π during the whole cycle of events, θ/ω being the steady-state transit time. Thus we do not have to regard the critical plane as moving backwards and forwards in the ultradynamic condition, for the critical plane coincides with the anode during the whole cycle of events. This is, of course, in accordance with the original definition; but as it appeared in the section dealing only with the steady state, some further definition for the ultradynamic state might have been necessary. Actually we see that it is not, the position of the critical plane as defined in the steady state being preserved in the ultradynamic condition. We have thus achieved what we set out to do, namely to determine our ultradynamic solution at a fixed plane.

As to what actually happens to individual electrons in the critical condition, it is clear that we cannot investigate this point by reference to the solution for individual electrons as given at (*b*) in (2.42), since it is not permissible to take $\omega(t-t_0)$ as 2π to correspond with the critical condition.

Associated with the critical condition is the obvious conclusion that, during that half-cycle which corresponds to an anode potential below the mean value, electrons would, at low frequencies, fail to reach the anode at all, except for voltage-amplitudes comparable with the initial velocity U_0 expressed in volts. This conclusion is satisfactory in the case of negligible space charge, but the existence of space charge concentrated near the anode may alter the situation. To investigate this point, we may write down the electron current I_c at the critical plane as follows: I_c

$$I_c = P_c U_0 + \rho_c U_0 + u_c P_c + \rho_c u_c,$$

where P_c is the steady-state value of space-charge density at the critical plane, and ρ_c is the fluctuating component of space-charge density at the critical plane. Now electrons will reach the anode at all instants of the negative half cycle of u_c if I_c has the same sign as $(P+\rho_c)$, which requires that P_c ρ_c

$$U_0 > -u_c.$$

This shows that the conclusion still applies when space charges are present. This does not prevent u_c from being greater than U_0 during the positive half cycle, but equation (120) can then only be applied during the positive half cycle and that part of the negative half cycle for which the condition $|u_c/U_0| < 1$ is satisfied.

With the above reservations, figure 4 gives the alternating velocity u_c in arbitrary units for values of p/ω between 0 and 2, for the case where $X_0/U_0H_c = 4$, which corresponds closely enough to the values of X_0, U_0 and H_c used in the example standardized in previous sections and also for the case $X_0 = 0$ (dotted). In figure 4 is shown only the coefficient of $(-\sin pt)$, designated $(-\hat{u}_i)$. In figure 5 is given the value of $(-\hat{u}_i)$ corresponding to negligible space charge $(\theta_c = \pi)$.

(7·3) *Alternating potential at critical plane.* Now the following equation connects the alternating potential v_c at the critical plane with the alternating velocity existing there: v_c

$$\frac{-e}{m} v_c = U_0 u_c + \int_{\theta=0}^{\theta=2\pi} \frac{\partial u}{\partial t} \frac{dx}{d\theta}\, d\theta.$$

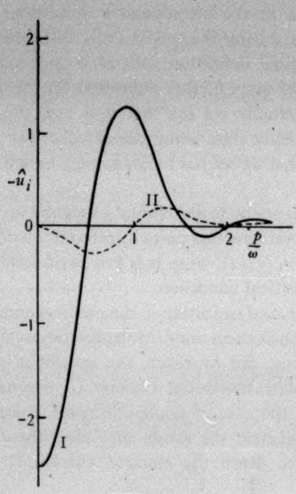


Figure 4.

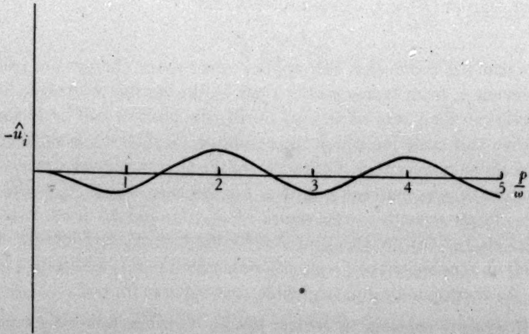


Figure 5.

Figure 4 may evidently be regarded as referring also to the alternating potential for values of p/ω which make the above integral of negligible value in comparison with $U_0 u_c$. The integral has not yet been evaluated, except in the case where U_0, X_0 are zero. In the absence of definite information regarding the importance of the integral term, figures 4 and 5 will be treated provisionally as referring to the alternating potential, subject to revision at some future date. The alternating potential will cut off in general during a part of the negative half-cycle, so that the dynamic condition has a distinct resemblance to the condition obtaining in an ordinary triode regenerative circuit. For $p/\omega = 1$ the peak value of v_c is

$$\frac{4\pi^2 m^2 c^2 X_0}{e^2 H_c^3} j_1.$$

Taking X_0 as 114 V./cm., H_c as 131·6 G., $-j_1$ as 1 mA. or 10^{-4} e.m.u., we obtain $v_c = -0{\cdot}57$ V. The above condition corresponds to a negative conductance of the order of 2 mA./V.

(7.4) *Oscillatory properties of planar magnetron.* Inspection of figure 4 reveals that the coefficient of $(-\sin pt)$ has positive maxima when

$$p=\omega;\quad p=2{\cdot}3\omega.$$

The relative insignificance of the maximum at $p = 2{\cdot}3\omega$ is in accordance with expectations (q.v.). The range of possible oscillation indicated by the positive values of $(-\sin pt)$ around the value $p/\omega = 1$ is as follows, when T_0 is written for $2\pi/p$ T_0

$$0{\cdot}625 < T_0/T < 1{\cdot}85 \qquad \text{......(121).}$$

The oscillations predicted by the range (equation 121) depend on the importance of electron-inertia and probably correspond to the characteristic oscillations of a solid-anode magnetron. The rapid clamping of the curve of figure 4 precludes the likelihood of obtaining any oscillations of higher frequency, the counterpart of which is well known in connection with the retarding field triode*. This result is in accordance with observation. On the other hand, figure 5 (for negligible space charge) *would* predict ranges of higher frequency oscillation. Oscillations with a cylindrical magnetron were produced by Zacek, originally in the year 1924†. An excellent summary of the present position of our experimental knowledge of magnetron oscillations is provided by Megaw‡. The corresponding oscillations in the retarded-field triode are the original Barkhausen-Kurz oscillations§. These have been extensively studied by workers too numerous to mention. Excellent summaries are to be found in papers by Hollman‖, Pierret¶, and Megaw**.

---

* Scheibe, *Ann. Phys.*, Lpz., **73**, 54 (1924); Potapenko, *Phys. Rev.* **39**, 625 (1932); Hollmann, *Z. f. Hochfrequenztechnik*, **37**, 145 (1931); Benham, *Electrical Communications XI*, **39**, 223 (1933).

† *Z. f. Hochfrequenztechnik*, **32**, 172 (1928).

‡ *Loc. cit.*

§ *Phys. Z.* **21**, 1 (1920).

‖ *Hochfrequenztech. u. Elektroakust.* **44**, 2, August 1934.

¶ *L'Onde Electrique*, **8**, 373 (1929).

** *J. Instn elect. Engrs*, **72**, 313 (1933).

(7.5) *The effect of cathode-temperature on oscillation-intensity.* The existence of an optimum in the relation between cathode-temperature and oscillation-intensity has been known since 1920*, but no satisfactory explanation has to my knowledge been proposed. The present paper contains all the material necessary for the establishment of a relation between cathode-temperature and the high-frequency energy developed by a planar magnetron diode in the critical condition. Comparison of theoretical results with experiment will constitute as severe a test as any of the fundamental correctness of the analysis.

In order to obtain the desired relation we shall first of all require to know the relation between cathode-temperature and total emission. It has been found that Richardson's law† holds quite accurately for any one coated cathode, provided it has been properly formed and aged. Measurements on several oxide-coated cathodes have led to the following average value of Richardson's constant b *b*

$$b = 1{\cdot}42 \times 10^4 \ {}^\circ\text{K}.$$

For a total emission of 200 mA./cm.² at a value 1173° K. of T_c the corresponding value of a in the simplified Richardson equation $J_s = ae^{-b/T_c}$, with J_s in mA./cm.², is given by

$$\log_{10} a = 14{\cdot}4.$$

With the above information and with the assistance of (4.5) and (4.6) we obtain the results in table 4 for J_s, U_0 and X_0 as functions of T_c, the anode current being supposed constant at 50 mA./cm.²

Table 4

$J = 50$ mA./cm.²

T_c (° K.)	J_s (mA./cm.²)	$\dfrac{J_s}{J}$	$-V_m$ (V.)	$10^3 x_m$ (cm.)	$10^{10} T_m$ (sec.)	X_0 (V./cm.)	$10^{-7} U_0$ (cm./sec.)	$10^7 X_0 U_0^{-1}$ (V.-sec./cm.²)
1279	2000	40	0·407	2·426	3·030	171·4	3·86	44·4
1235	800	16	0·2953	2·225	2·700	152·6	3·23	47·2
1222	600	12	0·2618	2·150	2·580	146	3·05	47·9
1190	300	6	0·1838	1·920	2·165	122·5	2·446	50·1
1173	200	4	0·1402	1·775	2·015	114·0	2·233	51·1
1145	100	2	0·0685	1·330	1·498	84·8	1·637	51·8
1132	75	1·5	0·0396	1·056	1·17	66·2	1·360	48·7
1125	60	1·2	0·0178	0·733	0·810	45·8	1·126	40·7
1118	50	1	0	0	0	0	—	0

The last column of table 4 gives the ratio X_0/U_0 which from table 3 is seen to constitute an important factor in determining the value of u_c. In particular, we note that when $p = \omega$ the coefficient of $(-\sin pt)$ is directly proportional to X_0/U_0. For a given value of β which is determined by j_1 we may say that the effect of cathode-temperature on the oscillation-intensity is to be measured by the ratio X_0/U_0. This ratio is seen from figure 6 to be a maximum when $T_c = 1145$° K. In practice the optimum will occur below or above this temperature according to whether the emission-density when $T_c = 1173°$ K. is above or below the value here assumed, e.g. 200 mA./cm.^{2}

---

* Barkhausen and Kurz, *loc. cit.*

† *Loc. cit.*

The form of curve shown in figure 6 is quite characteristic of experimental behaviour, though in some classes of oscillation* a maximum at one temperature has been found to be followed by a minimum at some higher temperature. No trace of this minimum has been discovered theoretically by working with still higher emission-densities, but the reason for this is not difficult to see. Table 4 is adjusted to a fixed anode current of 50 mA., whereas, in practice, if the cathode-temperature be progressively raised the anode current will continually rise. This applies in the case of coated cathodes for values of total emission up to many times the value of space current taken. This effect is well known among valve engineers as "poor saturation", though this term is also frequently applied to the other end of the scale (voltage saturation) in cases where the lack of a flat top to the anode characteristic is observed. Poor (space-charge) saturation is less noticeable with bright tungsten, but is still very noticeable when the tungsten filaments are external to the anode, as in McPetrie's case. The output curve as given by McPetrie corresponds to an anode current continually increasing as the temperature of the cathode is raised. A continually increasing value of J would be accompanied by a continually increasing value of j_1 which would *per se* tend to increase the oscillatory output as the temperature is raised.

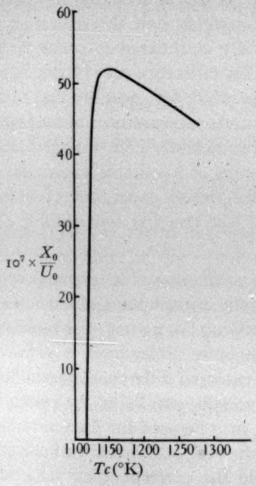


Figure 6.

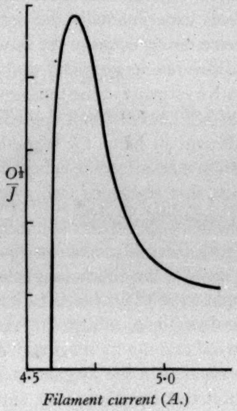


Figure 7.

---

* See McPetrie, *Wireless Engr*, p. 121 (1934).

O

Now McPetrie's curves show output as a function of filament current. The high-frequency voltage developed will be proportional to the square root of the output power O. If we divide this by the anode current we shall obtain a curve suitable for comparison with figure 6. The quantity $O^{\frac{1}{2}}/J$ is shown roughly on figure 7, and it is seen that there is no minimum following the maximum. In view of the marked difference in geometry and operating temperature and other differences, further information is required in order to establish quantitative agreement between theory and experiment.

In so far as theory predicts an optimum operating-temperature, the agreement with experiment may be regarded as promising. It will be seen that the theoretical optimum corresponds to an anode current of roughly half the saturation value, whereas experimentally the optimum is generally considered to occur for anode currents nearly equal to the saturation value. This difference may be due in part to the difference in geometry and operating-temperature but possibly also to a conservative estimate of the total emission. The accurate measurement of total emission requires a circuit which applies to the valve a sufficiently high voltage for a time insufficient to harm the cathode. The description of a suitable circuit for total-emission measurements is beyond the scope of the present paper, but it is of interest to note that emissions well over 2000 mA./cm.^{2} (see the first row of table 4) have been measured directly.

(7.6) *General conclusions concerning electron oscillations.* The expression (equation 120) for the alternating velocity existing at the critical plane of a magnetron is thought to be of fundamental importance as typifying the nature of virtual cathodes generally. Thus, whilst the virtual cathode as now understood is obtained by electrical and not by magnetic means, the only essential difference appears to lie in the existence in the magnetron case of a drift velocity parallel to the plates. It is possible that equation (120), suitably modified, may be used for the retarding-field triode. If T denotes the transit time between the real and virtual cathode of a retarding-field triode subject to no magnetic field the corresponding value of u_c is likely to be given by an expression of the type

$$u_c=\frac{\beta}{p^2}\left[\left(-1+\cos pT-\frac{eX_0}{mpU_0}\sin pT\right)\sin pt\right.$$
$$\left.-\left\{\sin pT+\frac{eX_0}{mpU_0}\frac{pT}{2\pi}(1-\cos pT)\right\}\cos pt\right] \quad \ldots\ldots(122),$$

which differs from equation (120) only in that we have replaced $(p^2-\omega^2)$ by p^2 and $2\pi/\omega$ by T. Actually equation (122) as it stands does not remain finite when $p=0$.

L. Tonks* was the first to draw attention to the important part played by the formation of a virtual cathode in the retarding-field triode, and proposed some sort of combination of his theory for the steady state with the dynamic theory of Gill and Morrell†. The present paper is directed largely towards the fulfilment of the requirements postulated by Tonks, and it is believed that the analysis here given for the plane magnetron is accurate as far as it goes. The cylindrical case is likely to present grave difficulties judging by preliminary analysis, and it is to be hoped that many of the conclusions arrived at as a result of the study of a plane geometry will not be substantially modified when the results for cylinders are eventually derived.

---

* *Loc. cit.*

† *Phil. Mag.* **44**, 161 (1922).

§ 8. ACKNOWLEDGMENTS

In conclusion I should like to express my indebtedness to Marconi's Wireless Telegraph Company Limited for facilities in connection with the preparation of this paper, which is published by their permission.

APPENDIX. CHANGE OF COORDINATES

The theory as given in two earlier papers* involves the solution of the equation

$$\left(\frac{\partial}{\partial t}+U\frac{\partial}{\partial x}\right)\left(\frac{\partial U}{\partial t}+U\frac{\partial U}{\partial x}\right)=\frac{4\pi c^2 e}{m}J \qquad \ldots\ldots(1).$$

Equation (1) is the electrical counterpart of Euler's form in which x and t are independent variables of the equation of motion for fluids moving in streamlines†. If we wish to employ coordinates which specify the position of an individual electron in the course of its passage between cathode and anode, we shall need to specify that the distance travelled by a chosen electron is dependent on the time (measured from some arbitrary zero) and also on the instant at which the electron in question started from the cathode.

x, t

Taking t' and t_0 as Lagrangian variables, we shall have

t', t_0

$$x=\phi(t', t_0),$$

$$t=t',$$

i.e. the electron which started at time t_0 (which $<t'$) has reached the plane x at time t'.

Let f be any function expressed in Eulerian coordinates, and $\bar{f}$ the same function when expressed in Lagrangian coordinates, i.e.

$f, \bar{f}$

$$f(x, t)=\bar{f}(t', t_0).$$

Then

$$\frac{\partial f}{\partial t}\,\delta t+\frac{\partial f}{\partial x}\,\delta x=\frac{\partial \bar{f}}{\partial t'}\,\delta t'+\frac{\partial \bar{f}}{\partial t_0}\,\delta t_0,$$

or

$$\frac{\partial f}{\partial t}\,\delta t+\frac{\partial f}{\partial x}\left(\frac{\partial \phi}{\partial t'}\,\delta t'+\frac{\partial \phi}{\partial t_0}\,\delta t_0\right)=\frac{\partial \bar{f}}{\partial t'}\,\delta t'+\frac{\partial \bar{f}}{\partial t_0}\,\delta t_0,$$

and, since

$$\delta t'=\delta t, \quad U=\frac{\partial \phi}{\partial t'},$$

---

* Part I: *Phil. Mag.* **5**, 641–62 (1928); Part II: *Phil. Mag.* **11**, 457–517 (1931).

† See Lamb, *Hydrodynamics*, p. 2 (fifth edition).

we have, comparing coefficients of ∂t, ∂t_0 respectively,

$$\left.\begin{aligned}\frac{\partial f}{\partial t}+U\frac{\partial f}{\partial x}&=\frac{\partial f}{\partial t'}\\ \frac{\partial \phi}{\partial t_0}\cdot\frac{\partial f}{\partial x}&=\frac{\partial f}{\partial t_0}\end{aligned}\right\} \qquad \text{......(2).}$$

Equations (2) are the formulae for change of variable, and show that, in the Lagrangian scheme, equation (1) simplifies, if we now write $t'=t$, to

$$\frac{\partial^3 x}{\partial t^3}=\frac{4\pi c^2 e}{m}J,$$

in which the independent variables are t and t_0 though the latter does not appear explicitly.

Writing $$\frac{4\pi c^2 e}{m}J=B+\beta\sin pt,$$

and integrating with respect to t, while t_0 is constant, we have (assuming the electron starts with zero velocity and acceleration)

$$\frac{\partial^2 x}{\partial t^2}=B\,(t-t_0)+\frac{\beta}{p}(\cos pt_0-\cos pt),$$

$$U=\frac{\partial x}{\partial t}=\frac{B}{2}(t-t_0)^2+\chi \qquad \text{......(3),}$$

$$x=\frac{B}{6}(t-t_0)^3+\psi \qquad \text{......(4),}$$

in which we have abbreviated by writing

χ $$\chi=\frac{b}{p}(t-t_0)\cos pt_0+\frac{b}{p^2}(\sin pt_0-\sin pt),$$

ψ and $$\psi=\frac{b}{2p^3}[p^2(t-t_0)^2\cos pt_0+2p\,(t-t_0)\sin pt_0+2\,(\cos pt-\cos pt_0)].$$

Having obtained a solution for U (equal to $\partial x/\partial t$) which is very simply arrived at, we seek to express the solution in the original coordinates in order to enable us to calculate physically interesting quantities, such as conductance, admittance, etc. From equation (3) we obtain

$$(t-t_0)=\left[\frac{6\,(x-\psi)}{B}\right]^{\frac{1}{3}},$$

$$\therefore\ \frac{B}{2}(t-t_0)^2=\frac{B}{2}\left(\frac{6}{B}\right)^{\frac{2}{3}}(x-\psi)^{\frac{2}{3}}$$

$$=U_1\left(1-\frac{\psi}{x}\right)^{\frac{2}{3}},$$

where we have written U_1 for the value of $\partial x/\partial t$ in the steady state ($\psi=0$).

δT Writing U instead of $\partial x/\partial t$, and $(T+\delta T)$ instead of $(t-t_0)$, where δT is to be regarded as the deviation from the mean or steady state value of transit time, we obtain, from (2) and (4),

$$U = U_1\left(1-\frac{\psi}{x}\right)^{\frac{2}{3}} + \bar{\chi}(t', t_0)$$

$$= U_1\left(1-\frac{\psi}{x}\right)^{\frac{2}{3}} + \chi(T, t) + \text{terms in } \delta T.$$

It is readily shown that the fundamental component of U contains no term in δT, while the sole contribution to the fundamental arising from the term in ψ is

$$-\frac{2U_1}{3x}\psi.$$

Writing

$$U_1 x^{-1} = 3T^{-1},$$

we obtain, for the fundamental u,

$$u = -2T^{-1}\psi + \chi.$$

Giving χ and ψ their values, we have

$$u = \frac{\beta}{p^2}\left[-\sin pt - \sin pt_0 + \frac{2}{pT}\cos pt_0 - \frac{2}{pT}\sin pt_0\right],$$

in which

$$t_0 \equiv t - T,$$

in agreement with the solution given on p. 648 of part I.

Let us now derive the potential in Lagrangian coordinates. In Eulerian coordinates we have

$$-\frac{e}{m}\frac{\partial V}{\partial x} = \frac{\partial U}{\partial t} + U\frac{\partial U}{\partial x}.$$

In Lagrangian coordinates t, t_0 we have, using equations (2),

$$-\frac{e}{m}\frac{\partial V}{\partial t_0} = \frac{\partial x}{\partial t_0}\cdot\frac{\partial^2 x}{\partial t^2};$$

$\partial x/\partial t_0$ may be obtained from (3), thus

$$\frac{\partial x}{\partial t_0} = -\frac{B}{2}(t-t_0)^2(B+\beta\sin pt_0).$$

Hence

$$\frac{\partial x}{\partial t_0}\cdot\frac{\partial^2 x}{\partial t^2} = -\frac{B^2}{2}(t-t_0)^3 - \frac{B\beta}{2}(t-t_0)^3\sin pt_0 - \frac{B\beta}{2}\frac{(t-t_0)^2}{p}(\cos pt_0 - \cos pt)$$
$$-\frac{B^2}{2}\frac{(t-t_0)^2}{p}(\sin pt_0\cos pt_0 - \sin pt_0\cos pt).$$

This equation shows incidentally that harmonics are present in the potential if the current through the system is free from harmonics. Interesting ourselves only with the fundamental, we shall omit the last term. Integration with respect to t_0 yields the result

$$\frac{-e}{m}V = \frac{B^2T^4}{8} + \frac{B\beta}{2p^4}[p^3T^3\cos pt_0 + 2p^2T^2\sin pt_0 - 4pT\cos pt_0$$
$$-\tfrac{1}{3}p^3T^3\cos pt + 4(\sin pt - \sin pt_0)],$$

where $T \equiv t - t_0$.

The above equation may be reconverted to Eulerian coordinates by a process similar to that adopted in the case of the velocity, and the result agrees with the solution obtained by working in Eulerian coordinates throughout (see p. 510 of part II).

Turning now to the cylindrical case, equations (41) and (42) on p. 501 of part II,

$$\frac{\partial U}{\partial t}+U\frac{\partial U}{\partial r}=\frac{-e}{m}\frac{\partial V}{\partial r},$$

$$\left(\frac{\partial}{\partial t}+U\frac{\partial}{\partial r}\right)\left[r\left(\frac{\partial U}{\partial t}+U\frac{\partial U}{\partial r}\right)\right]=\frac{2ec^2}{m}J.$$

These become, as before,

$$\frac{\partial r}{\partial t_0}\cdot\frac{\partial^2 r}{\partial t^2}=\frac{-e}{m}\frac{\partial V}{\partial t_0},$$

$$\frac{\partial}{\partial t}\left(r\frac{\partial^2 r}{\partial t^2}\right)=\frac{2ec^2}{m}J.$$

The last equation requires the integration of an equation of the type

$$r\frac{\partial^2 r}{\partial t^2}=f(t).$$

The difficulties in connection with such an integration are such as to render a treatment in Lagrangian coordinates hardly less intractable than the treatment already given in Eulerian coordinates on p. 501 of part II.

Summarizing the above, we see that by means of a suitable change of variable it is possible to avoid the necessity for solving tiresome differential equations. A simplification of the equations results also in the cylindrical case, which, however, does not appear to permit of exact solution in either system of coordinates. The above conclusions break down in more complicated cases. Thus if initial velocities and accelerations, or a magnetic field, be included, the process becomes intractable and the method may become too inaccurate to be of value.

DISCUSSION

Prof. C. L. Fortescue. What is the exact mechanism whereby a two-electrode valve can have amplifying properties which give it the nature of a negative resistance? And are there any experiments to prove the author's theory to be correct—can a parallel-plate diode oscillate? Llewellyn says that this is probably impossible.

Author's reply. I find Prof. Fortescue's first question a little difficult to understand. The negative-resistance property is hardly less fundamental than the electron inertia, which is responsible for phase-differences corresponding to a negative power factor under favourable conditions. The amplifying properties, if any, would in my opinion be incidental to the negative resistance, and depend also on the external circuit. The mechanism of electronic oscillations in diodes and triodes is dealt with in an article due to appear in the *Wireless Engineer*. With regard to experimental confirmation, Müller* has recently established the existence of oscillations in a carefully constructed low-loss full-wave parallel-plane diode. These oscillations lie within the zone predicted by my theory, which also predicts the correct intervals for dwarf waves in a triode oscillator†.

The above remarks all refer to diodes or triodes without magnetic field. So far oscillations in a parallel-plane diode with magnetic field have not to my knowledge been observed. Any experiments conducted with a view to obtaining such oscillations would be subject to the difficulty of obtaining a sufficiently uniform magnetic field throughout the space between the plane electrodes. If these electrodes are not large compared with their separation edge, effects occur which may mask the oscillations or which may result in spurious oscillations due to a proportion of electrons missing the anode and returning to it from behind.

* *Hochfrequenztech. u. Electroakust.* **43**, 6, 195–199 (1934).
† *Elec. Comm. XI*, **39**, 223 (1933).

537·743

THE ABSOLUTE MEASUREMENT OF ELECTRICAL RESISTANCE BY A NEW ROTATING-COIL METHOD

By H. R. NETTLETON, D.Sc., Lecturer in Physics, Birkbeck College
and E. G. BALLS, M.C., B.Sc., A.I.C., Birkbeck College

Received July 11, 1934. *Read November* 2, 1934.

ABSTRACT. A rotating coil of mean radius α lies symmetrically between two fixed twin coils of radius a. If the ratio α/a lies between 0·58 and 0·53 it is easy to arrange, by merely adjusting the distance between the twin coils, that the mutual inductance M between the rotating coil and the two fixed coils shall be very accurately proportional to the angle θ of displacement from the conjugate positions over a range of some 10° of arc on either side of the zeros. The constant K of the relation $M=K\theta$ can then be accurately measured by a method here described in which θ is deduced from a mutual-inductance ratio. At the same time errors in the calibration of the inductometer used are eliminated.

If such a coil spins with an angular velocity ω, while a current C traverses the fixed twin coils, a uniform e.m.f. ωCK can be drawn off a commutator on the rotating shaft and made to balance an e.m.f. CR drawn off an adjustable resistance R carrying the same current. Thus $R=\omega K$.

If now the twin coils are brought rather closer together, the law of variation of M with θ takes the form $M/\theta=K+A\theta^2-B\theta^4$ over a displacement of 35° of arc from the zeros, where A and B are very small positive constants and M/θ attains a maximum value K', where $\theta^2=A/2B$ and scarcely changes over a range of 3° in this neighbourhood. This allows larger sectors to be used on the commutator and the constant K' of the relationship $R=K'\omega$ can be accurately determined for a suitable commutator *in situ*.

The experimental work is mainly devoted to a study of the laws of inductance on which the method depends and to the determination of the constants K and K', but preliminary spin experiments are very hopeful and resistances of between 0·32 Ω. and 0·64 Ω. have been measured absolutely by means of commutators with sector contacts of 23° and 47° of arc. Owing to the relatively large e.m.f.'s involved, the method is very sensitive and a fluxmeter can be used as the balance-detector for hand-controlled stroboscopic spins. In other experiments a synchronized television motor was used.

§ 1. INTRODUCTION

THE well-known method of measuring electrical resistance absolutely with the aid of a spinning coil was originally suggested by Weber and put forward independently by Kelvin to the Electrical Standards Committee of the British Association in 1863. It is very fully described in the reports of the British Association covering the period 1862–67; the experiments were carried out principally by Maxwell, Stewart, and Jenkin. Later determinations by this method were undertaken by Rayleigh and Schuster*; Rayleigh† and H. Weber‡.

---

* *Proc. roy. Soc.* **32**, 104 (1881).

† *Phil. Trans.* **173**, 661 (1882).

‡ *Der Rotations-inductor* (Leipzig, Teubner, 1882).

The method involves fundamentally the measurement of a coil-area, a galvanometer constant, a speed of rotation, and the deflection of a needle at the centre of the spinning coil. Corrections are necessary for the moment of the magnetic needle, the torsion of the supporting fibre, and the self inductance of the coil. It is not surprising, therefore, that the method has largely given way to the methods of Lorentz and Campbell which involve quantities more easy to determine and are applicable to the measurement of external resistances with potential-leads.

In 1880 Carey Foster* suggested an interesting null method involving the same principle as the British Association method. In this arrangement a steady current is passed through a tangent galvanometer of known principal constant and through the resistance to be measured. The e.m.f. across the potential leads of this resistance is balanced against that derived from a coil spinning in the earth's field, the balancing circuit being completed through commutators only over some 20° of arc. The middle of the period of contact was made to coincide with the instant when maximum e.m.f. was induced in the spinning coil, and the extreme variations of the e.m.f. during contact was 1·83 per cent. Though this null method has the great advantage of dispensing with the corrections necessary in the original method and is applicable to the measurement of an external resistance, the same fundamental quantities are involved. The correction necessary for the angle of contact, which must be measured, depends for its validity on a symmetrical setting of the commutators about the position of maximum e.m.f. which is difficult to locate with precision, and thermoelectric effects are likely to be troublesome.

The object of the present communication is to describe a preliminary investigation of a sensitive null method by which a resistance may be measured absolutely in terms of a mutual inductance and a frequency.

§2. THE THEORY OF THE METHOD

Simple form of the method. A coil spinning uniformly about a horizontal diameter lies between two larger fixed twin field coils having their planes horizontal and so separated in the theoretically simplest type of experiment that the mutual inductance M between the rotating and fixed coils in series and conjunction is, over a range of some 10° of arc on either side of zero, very accurately proportional to the angle θ of displacement from the position of zero mutual inductance. The axis of rotation lies in the magnetic meridian and the earth's vertical flux through the rotating coil is neutralized by a small current passing through large compensating coils in the Helmholtz position.

M

θ

A steady current C of about one ampere is passed through the twin fixed coils and through a variable manganin resistance which is adjusted until the e.m.f. across its potential-leads is balanced by that across the commutating sectors of the rotating coil. These make contact through fixed brushes with a galvanometer over some 20° of arc, during which the e.m.f. arising from the uniform spin is constant.

C

---

* *B.A. Reports*, p. 426 (1881).

M

Since the mutual inductance M at any position θ over the range of contact is given by the linear relationship

$$M = K\theta \qquad \ldots\ldots(1),$$

K

where K is constant, we have for the flux F through the spinning coil at any position θ

C

$$F = CM = CK\theta \qquad \ldots\ldots(2),$$

and for the numerical value of the e.m.f.

E, t

$$E = dF/dt = CK\omega \qquad \ldots\ldots(3),$$

ω
R

where ω is the constant angular velocity of rotation. Whence, equating this to the e.m.f. CR across the balancing resistance R, we have

$$R = K\omega = 2\pi nK \qquad \ldots\ldots(4),$$

n

where n is the number of revolutions per second.

In view of the remarkable accuracy of the linear law connecting M and θ over considerable range the constant K can be determined with precision, and a method of doing this is described below which not only avoids the direct measurement of angles but at the same time eliminates errors in the stud-calibration of the mutual inductometer used. Further, on account of the linear law which gives rise to a uniform e.m.f. the exact angle of contact is not required and, although symmetry of contact is readily obtainable by means of inductance measurements, slight departure from symmetry is unimportant. Thermoelectric effects are rendered insignificant by adjusting the earth-coil current for zero galvanometer deflection when the current C is broken and the resistance R is connected across the contacts of the rotating coil spinning at the frequency n. On application of the current C relatively large opposing e.m.f's. come into action, giving high sensitivity, and on reversal of C throughout the balance is preserved.

Extension of the method to large angles of contact. So far we have supposed that the twin field coils are so separated that the linear law $M = K\theta$ is rigorous and that an unvarying e.m.f. $CK\omega$ is drawn off the rotating coil over an angle limited to some 20° of arc. Let us now consider a more general case applicable when the commutator sectors are enlarged to some 50° of arc.

M'
θ'
M''
θ''

Let M' be the change in the mutual inductance between the rotating coil and the twin fixed coils over the angle of contact θ' of the sectors with the brushes in the region of the first conjugate position, and let M'' be the change in mutual inductance over the corresponding angle of contact θ'' around the second conjugate position. Then the flux changes CM', CM'' occur in times θ'/ω, θ''/ω respectively. Thus the average e.m.f. during the contacts of one revolution is $C\omega\,(M'/\theta' + M''/\theta'')/2$; and if this is balanced on a ballistic galvanometer or fluxmeter by the opposing e.m.f. CR, drawn off the adjustable resistance R, we have:

$$R = \omega\frac{1}{2}\left(\frac{M'}{\theta'} + \frac{M''}{\theta''}\right) = \omega K' \qquad \ldots\ldots(5).$$

Now while in general it would not be possible to measure precisely the sweeps of inductance M' and M'' over precise sector angles θ' and θ'', it is easily possible to measure with great precision the ratios M'/θ' and M''/θ'' provided that M/θ is rendered constant in the neighbourhoods of the sector-extremities. This, it has been found, can be readily accomplished by bringing the twin field coils rather closer together than when they are set to yield the linear law, so that their mutual inductance for a range of some 35° of arc on both sides of the zero obeys a law of the form

$$M/\theta = K + A\theta^2 - B\theta^4 \qquad \text{......(6).}$$

It will be seen below that the coils were so separated that

$$M/\theta = 124{\cdot}932 + 3{\cdot}665 \times 10^{-4}\theta^2 - 3{\cdot}05 \times 10^{-7}\theta^4 \qquad \text{......(7),}$$

where M is in nominal microhenries and θ is in degrees of arc. This gives a maximum value for M/θ of 125·042 μH. per degree at 24°·5. The value 125·040 corresponds to both 22°·8 and 26°·1, so that over this range of 3°·3 of arc M/θ may be regarded as constant and equal to $dM/d\theta$.

By making one sector larger than the other so that the range of contact is determined by the smaller sector, θ' and θ'' were made equal at a value of about 46°·8. M'/θ' and M''/θ'', rendered almost equal by symmetrical setting, were measured by observations of M and θ in the neighbourhoods of the sector edges, readings being taken when the sectors were (*a*) just on and (*b*) just off the contact brushes, the differences in angle between the on and the off readings being narrowed down to about 0°·1. Under such circumstances, the constant K' can be measured with high precision.

It should be observed that although in this method of working with a large angle of contact the e.m.f. drawn off the rotating coil is not quite constant, the extreme variation of e.m.f. is only 0·16 per cent over a contact of 47° as against 1·83 per cent over 22° of arc in Carey Foster's method. Moreover in this method the e.m.f. at the point of leaving the sectors is equal to the average e.m.f. over the whole contact.

Design and arrangement of coils. The design and arrangement of coils which provide the simple inductance laws stated above are based on the following theory. If two concentric circles have the ratio of their radii α/a equal to 0·506078* the mutual inductance between them is so accurately proportional to the angle of displacement from the conjugate positions that the rising deviation from a straight-line law amounts to only 4·2 parts in a million at 7° of displacement and 18 parts in a million at 10° of displacement. With a slightly larger ratio of α/a, the deviation may be distributed so that over a range of 7° it never exceeds 0·75 parts in a million. On the other hand, if a coil of radius α lies between two circles of radius a separated by a distance $2x$, the ratio α/a must be increased to bring about a similar approach to linearity. Thus if $x/a = 0{\cdot}19438$, the ratio α/a must be raised to 0·5461 to preserve the limiting linear law, the deviation at 7° being now 5·1 parts in a million. This deviation may likewise be distributed and diminished by slightly raising the ratio α/a.

The practical significance of this theory, coupled with the fact that the primary effect of multiplicity of layers is to alter slightly the effective radii of the coils, lies in the result that if α/a lies between the limits $0{\cdot}58 > \alpha/a > 0{\cdot}52$, linearity within the accuracy of experimental measurements may be secured over some 12° with multiple-layered coils by merely adjusting the distance of separation between the larger twin coils. Such twin coils may at any time be joined in opposition, thus enabling the smaller coil to be set symmetrically.

---

* Nettleton and Llewellyn, *Proc. phys. Soc.* **44**, 195 (1932).

In general we may express the mutual inductance between twin coils and a smaller coil displaced from the conjugate positions by an angle θ by a series of the type

$$M = K\theta + A\theta^3 + B\theta^5 + \ldots\ldots \qquad \ldots\ldots(8),$$

where the constants K, A, B etc. depend on α, a, x and the number of turns, and may be evaluated with the aid of Legendre functions, though the process is laborious.

If α/a lies between the limits given, the coefficient A may be rendered zero by adjusting the separation $2x$. The succeeding coefficients are then small and negative, and a limiting linear law with accuracy of the order already stated results. If now the separation $2x$ is reduced, A assumes a small positive value and we have in practice with high accuracy over 30° of arc

$$M = K\theta + A\theta^3 - B\theta^5 \qquad \ldots\ldots(9),$$

which gives a maximum value of $M/\theta = K + A^2/4B$ at positions given by $\theta^2 = A/2B$. These positions are those proposed for the sector edges and define the ideal angle of contact for the new and closer distance of separation.

The e.m.f. on uniform rotation is everywhere proportional to $dM/d\theta$. Thus the minimum e.m.f. over the sector contacts is represented by K when $\theta = 0$, the maximum e.m.f. by $K + 9A^2/20B$ when $\theta^2 = 3A/10B$, and the average e.m.f. by $K + A^2/4B$ (which is the actual e.m.f. when $\theta^2 = A/10B$ or at the sector edges) where $\theta^2 = A/2B$.

Measurement of angle by a mutual inductance method. This method of measuring an angle of displacement is particularly suitable for the present purpose and is based upon the fact that the mutual inductance between a small solenoid of designed dimensions and two large twin coils between which it rotates can be rendered with great accuracy proportional to the sine of the angle of displacement from the conjugate positions.

L

The mutual inductance between twin circles A and B, figure 1, and a single layered solenoid C of radius α and length $2L$, rotated from the conjugate position by an angle θ, is given by the expression

M_θ

$$M_\theta = Gq\sum_{n=1}^{n=\infty} \frac{2}{n(n+1)(n+2)} \left(\frac{s}{r}\right)^{n-1} P_n'(\cos\psi)\,\frac{P_{n+1}(\cos\phi)}{\cos\phi}\,P_n(\sin\theta) \;\ldots\ldots(10),$$

P_n, P_n'

G

q, s, r, ϕ, ψ

where P_n is the Legendre function of the first kind, of order n; P_n' its differential coefficient; n an odd positive integer, G the galvanometer constant of A and B together at the origin of symmetry; q the total area of C; and s, r, ϕ, and ψ are sufficiently defined by the figure.

If the angle solenoid C is of small radius, this series converges rapidly and only the first three terms are of importance. They lead to the expression

$$M_\theta/Gq = \sin\theta + K_1 P_3(\sin\theta) + K_2 P_5(\sin\theta) \qquad \ldots\ldots(11),$$

where

$$K_1 = \frac{3\cdot 4}{2}\cdot\frac{1}{r^4}\left[x^2 - \frac{a^2}{4}\right]\left[\frac{L^2}{3} - \frac{\alpha^2}{4}\right]$$

and

$$K_2 = \frac{5\cdot 6}{2}\cdot\frac{1}{r^8}\left[x^4 - \frac{3}{2}x^2a^2 + \frac{a^4}{8}\right]\left[\frac{L^4}{5} - \frac{L^2\alpha^2}{2} + \frac{\alpha^4}{8}\right],$$

or when the result is expressed in powers of $\sin\theta$

$$\begin{aligned} M_\theta/Gq = {} & \sin\theta\,[1 - 3K_1/2 + 15K_2/8] \\ & + \sin^3\theta\,[5K_1/2 - 35K_2/4] + \sin^5\theta\,[63K_2/8] \qquad \ldots\ldots(12). \end{aligned}$$

If α/a is less than $\frac{1}{4}$, and K_1 is rendered sensibly zero by making $L^2/3$ equal to $\alpha^2/4$ or x^2 equal to $a^2/4$, K_2 also is always small and a close approach to the sine law results. If the coils are multiple-layered to the extent used in this research, this effect may be treated by a method due to Maxwell* and shown to be negligibly small.

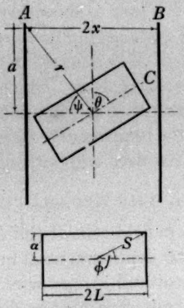


Figure 1.

In the angle solenoid used in our experiments $2L = 3{\cdot}78$ cm. and $\alpha = 2{\cdot}18$ cm. as found with the aid of a standard solenoid. The mean radius a of the twin coils was 16·41 cm. and in the least favorable position used, namely when the coils were closest and $\psi = 78°$, these figures give $K_1 = -1{\cdot}09\times10^{-5}$ and $K_2 = -3{\cdot}20\times10^{-5}$. Whence if $M_{\max.}$ denotes the value of M when $\theta = 90°$

$M_{\max.}$

$$\sin\theta = \frac{M_\theta}{M_{\max.}}\,[1 + 7\times10^{-7} - 2{\cdot}5\times10^{-4}(\sin^2\theta - \sin^4\theta)] \qquad \ldots\ldots(13)$$

and the sine law is very accurate, but it should be observed that the value of K_1 is necessarily uncertain as errors in the measurements of L and α have large effects on its calculated value. Some experimental tests upon this angle coil have already been described by Llewellyn†, but in view of the special advantage of the method for the present purpose and its sensitivity to two or three seconds of arc over considerable range, further improvements in the method and in the tests of its accuracy are contemplated.

---

* *Proc. phys. Soc.* **42**, 507 (1930).

† Thesis, Ph.D. degree, University of London.

Method of eliminating local inductometer errors. Essentially the methods here described with either large or small angles of contact require the determination of a constant M/θ, the angle θ being measured from the relationship $\sin\theta = M_\theta/M_{\text{max}}$. All mutual inductances are measured on a Campbell inductometer. This inductometer is first carefully calibrated so that all the stud readings are known in terms of 100 divisions of the scale, so that inductances can be read in nominal microhenries, subject to calibration errors. Further, the rotating coil is wound with such a number of turns that its mutual inductance with the twin field coils is so close to the value of the mutual inductance between the angle coil and the same field coils that up to at least 25° of arc the readings are within 100 microhenries; thus the same thousands and hundreds studs are in use for both readings. With the widening gap between the radian law and the sine law, the difference of reading at 40° of arc is still only some 300 μH.

Under such circumstances it is easy to show that any errors of the order possible in the calibration are rendered quite negligible as regards the determination of the ratio M/M_θ, and the accuracy of M/θ is solely dependent on the accuracy of the inductometer at the angle-coil reading $M_{\text{max.}}$, which in our case was some 7000 μH.

In future work we shall aim at a maximum angle-coil reading of just over 10,000 μH., which will enable the fundamental dimensional length measurement to be checked by a 10 millihenry standard without any resort to the calibration curve.

§ 3. THE APPARATUS

The formers of the fixed twin field coils *A* and *B*, figure 2, were constructed of dexonite. The channels, of radial depth and axial breadth 3·6 cm., were each filled with 504 turns of double-silk-covered copper wire, s.w.g. 16, the mean diameter of winding being 32·8 cm. The coils, separated by distance pieces *D*, *D*, were locked together and firmly supported.

The rotating coil *C* consisted of 80 turns of double-silk-covered copper wire, s.w.g. 26, wound on a solid mahogany former, the channel having axial breadth 1 cm. and radial depth 0·3 cm. The mean diameter of winding was about 19·0 cm. A hole of diameter 5 cm. through the centre of the former served for the insertion of the angle solenoid when required. The hole was lengthened by attaching to each face of the former wooden rings, one of which is seen at *E*; to the ends of these rings were fastened small brass brackets with screw adjustments enabling the angle coil to be set centrally and secured in position.

The portions of the brass shaft to which the rotating coil *C* was attached terminated in pieces of channel brass *F*, *F* which held the coil by brass bolts. Some play within the channel pieces, which allowed room for packing, enabled the coil to be set symmetrically about the axis of rotation before the bolts were screwed up tightly. The brass shaft was of external diameter 1·27 cm. and had a central hole of bore 0·5 cm. which permitted bell flex leads from the rotating coil to be led through the portion *G* of the shaft to terminals *t*, *t* on the sectors of the commutator *M*.

The commutator M consisted of a cylindrical ebonite piece some 6·5 cm. in diameter and 3·3 cm. long, through the centre of which passed stout brass tubing which enabled it to be slipped over the shaft and fixed thereon by heavy screws. The brass sectors let into the ebonite were diametrically opposite. In the particular commutator shown inset in figure 2 on a larger scale, one sector S consisted of a brass piece tapering along its length from 44° to 51° of arc while the other sector was parallel along its length and of width equivalent to 60° of arc. The framework N carried spring brass brushes adjustable in width and height and provided with terminals. The fly-wheel L, the stroboscopic disc R, the flexible unions U and the well-separated motor T are seen in the diagram.

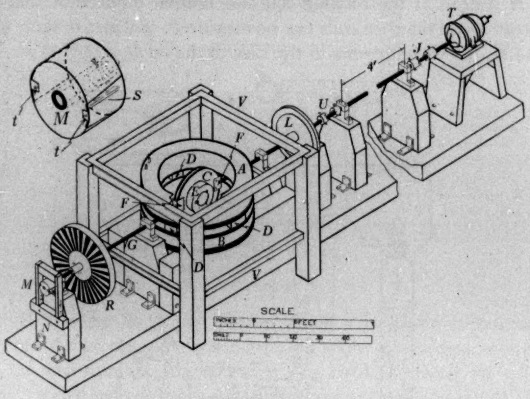


Figure 2.

The large horizontal coils V, V of sides 49 × 55 cm. are placed in the equivalent Helmholtz position and serve for the neutralization of the earth's vertical flux through the rotating coil when fed by a small current derived from accumulators.

The main electrical circuit will be readily understood from figure 3 with little description. The current of about 1 A., derived from accumulators, can be sent in either direction through the twin coils A and B, the quadrant key Q_3 serving to join them either in conjunction or in opposition but always in series. The same current passes through standardized resistances of 0·20029 Ω. and 0·50006 Ω. and through the variable resistance R to be adjusted and measured, all of which can be connected by potential leads with a good thermoelectric potentiometer which reads from 0 to 90 millivolts so that the value of R can be checked by comparison within a few minutes of measuring its resistance absolutely. The resistance R consisted of a box of manganin resistances of nominal values varying from 0·005 Ω. to 2·0 Ω. in series with a short semi-circular copper wire, of s.w.g. 18 provided with a movable potential contact. The e.m.f. across R can be adjusted to neutralize that across the brushes *b* of the uniformly spinning coil, balance being observed on a galvanometer or fluxmeter provided with a tapping key.

The circuit also readily permits of the measurement of the mutual inductance between the rotating coil, when stationary, and the twin field coils as well as of the measurement of the angle between them. For this purpose the primary of a Campbell mutual inductometer is switched in at the quadrant key Q_1 to be in series with the fixed coils. The rotating coil through the sector terminals *t*, *t* is thrown into series with the secondary of the mutual inductometer at the quadrant key Q_2 and the galvanometer is included in this loop and detached from *R* by rocking over the mercury switch *D*. Alternatively, the angle coil may be thrown into the secondary circuit at *H* instead of the rotating coil and mutual inductance, and hence the angle between it and the twin coils can be measured. Balance is easily obtained to within 0·1 μH., which represents in the case of the angle coil 3″ of arc.

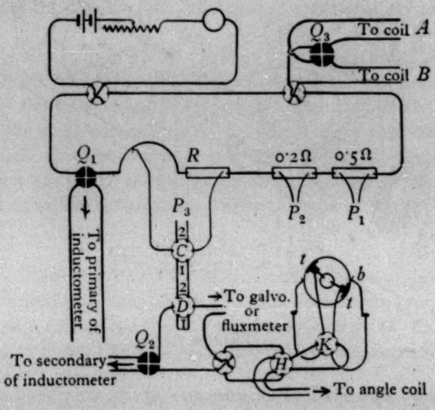


Figure 3.

The most constant speed of revolution was obtained by using a synchronized television motor, the cogged wheel of which had thirty narrow teeth separated by gaps four times the width of a tooth. The synchronizing impulse is fed to coils actuating an electromagnet pulling upon the teeth of the cogged wheel. The impulses used were derived from valve-maintained König tuning-forks which had been calibrated against one another by the method of visible beating[*]. In figure 4 the complete valve circuit is shown for maintaining the fork, synchronizing the motor and supplying a neon lamp with the tuning-fork frequency for viewing the stroboscopic disc *R*, figure 2, now provided with thirty lines. Because the television motor was rather weak in power, it was somewhat overrun and required careful rheostat adjustment to maintain synchronization for brief periods, but perfect balances were obtained on the galvanometer at three different speeds, viz. 8·533, 10·667 and 12·800 revolutions per second corresponding respectively to forks of frequencies 256, 320, and 384 vibrations per second.

---

[*] Nettleton and Balls, *Proc. phys. Soc.* **45**, 545 (1933).

For other speeds a stronger motor was used in conjunction with other stroboscopic disc rulings, viewed by means of the fork-controlled neon lamp, and the average speed was maintained as constant as possible with rheostats and hand-friction control. Though the imperfections of this control were manifested on a galvanometer by the oscillations of the spot of light, excellent balance was obtained by means of the following artifice. A Grassot fluxmeter of the silk-fibre-suspension type* shunted by some 80 Ω. was substituted by a rock-over switch for the galvanometer, and the tapping-key was depressed while the average known speed of revolution was maintained. Any advance of a stroboscopically viewed line was immediately rectified by increasing the friction and enforcing its return. Departure from balance was then manifested by the growing drift of the fluxmeter pointer in one or the other direction, and was rectified by adjustment of *R* until, after a considerable run, the pointer maintained its zero value. This method was proved to be satisfactory by deliberately allowing lines to escape and return in a time small compared with the period of the fluxmeter. Moreover the sensibility was good, the balance of *R* being sharp on the slide wire, while even small departures from the correct current in the earth coils used for neutralizing the earth's vertical component could readily be detected by this fluxmeter method.

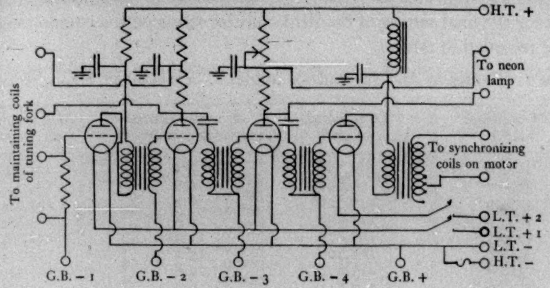


Figure 4.

§ 4. EXPERIMENTAL TESTS

Attainment of the linear law $M = K\theta$. The distance pieces between the twin field coils *A* and *B* were adjusted by trial, a few readings of the type given in table 1 sufficing until close approximation to the linear law was attained. Symmetry of setting was acquired by throwing *A* and *B* into opposition and moving them together until the mutual inductance between them and the rotating coil was as closely as possible zero in all positions. The angle coil likewise was set symmetrically by opposition tests, and its positions of zero mutual inductance with the fixed coils in conjunction were made to agree very closely with the corresponding conjugate positions of the rotating coil. Readings of the mutual inductance M between the rotating coil and the field coils, and the readings of the corresponding mutual inductance M_θ between the angle coil and the field coils, were then taken over all four quadrants at closely corresponding positions. The maximum angle-coil inductance was observed also and the angle θ in seconds of arc was then evaluated from the sine law for all positions. The corresponding values of M and θ in all four quadrants were added together and the ratio M/θ was obtained from $\Sigma M/\Sigma\theta$. The values obtained for the final setting of the field coils for spins with a commutator of small angle are recorded in table 1.

---

* The recent control-less jewel-pivoted type is less satisfactory owing to solid friction.

Table 1. Values of M/θ for various angles θ. $x/a = 0{\cdot}247$, $M_{\text{max.}} = 7085{\cdot}17$

ΣM nominal (μH.)	$\Sigma\theta$ (seconds of arc)	θ, approximate (degrees)	M/θ nominal (μH./degree)
2560·6	74,726	5·2	123·3$_{59}$
3279·2	95,698	6·6	123·3$_{58}$
4033·9	117,726	8·2	123·35$_{5}$
4833·2	141,050	9·8	123·35$_{7}$
5649·5	164,875	11·45	123·35$_{5}$
6448·8	188,209	13·1	123·35$_{1}$
7320·8	213,675	14·8	123·34$_{1}$
8148·6	237,844	16·5	123·33$_{7}$
8947·9	261,176	18·1	123·33$_{6}$
9781·3	285,502	19·8	123·33$_{6}$
10659·4	311,146	21·6	123·331
12450·2	363,614	25·25	123·265
15043·9	439,795	30·5	123·144

A plot of M/θ against θ reveals peculiarities due to imperfect symmetry, but the approach to linearity is remarkable and the value of K for contacts of approximately known angle can be readily determined with precision.

With the coils in this position the semi-angle of the contacts used in the spin experiments was 11°·6, and accordingly the value of the constant K was taken as 123·355 nominal microhenries per degree. The correcting factor to convert to true microhenries was found with the aid of a standard inductance to be 1·00035, and hence for this arrangement of coils and contacts the experimental value of K is 123·398 μH. per degree of arc.

Spin experiments with small angle of contact. The small commutator was readily adjusted systematically with the aid of the inductometer until the mutual inductance between the rotating coil and the fixed coils had approximately the same value, at the four boundaries of make and break between sectors and brushes, which are crossed during each revolution. The value in question was some 1425 μH., corresponding to a semi-angle of 11°·6. The current in the earth coils was then adjusted, when the rotating coil was spinning with the main current broken, until no deflection was obtained on the galvanometer when closed through R and the commutator of the revolving coil. The speed of revolution was then maintained at a constant value, determined by the tuning-fork and the stroboscopic disc in use, while R was adjusted until accurate balance was obtained on the galvanometer or fluxmeter when a main current of the order of an ampere traversed the circuit in either direction. The balancing value of R was then immediately checked on the thermoelectric potentiometer by comparing it with one of the standard resistances—usually the 0·50006-Ω. resistance. The value of R in c.g.s. units is given by the expression

$$R=36\times 123{\cdot}398\times 10^{4}\times n,$$

where n, the number of revolutions per second, is obtained by dividing the frequency of the fork by the number of lines on the stroboscopic disc. Table 2 gives the results obtained.

Table 2. Absolute measurement of resistance. $K=123{\cdot}398$ μH./degree

Frequency of fork (c./sec.)	Sectors on disc	n (rev./sec.)	$R\times 10^{-9}$ (c.g.s. units)	Resistance by comparison (Ω.)
256	30	8·533	0·37908	0·37889
320	30	10·667	0·47385	0·47365
384	30	12·800	0·56862	0·56842
384	36	10·667	0·47385	0·47386
512	36	14·222	0·63180	0·63193
384	40	9·600	0·42646	0·42647
512	40	12·800	0·56862	0·56879

The use of the synchronized television motor was limited of necessity to the first three experiments. Fluxmeter balances were taken in the other tests.

Adjustment of twin coils for use with large contacts. The twin coils A and B were now brought closer together and set symmetrically with respect to both the rotating coil and the angle coil by opposition inductometer tests as previously described. The variation of M and θ was then explored over all four quadrants by taking readings of M and M_θ in closely corresponding positions.

On adding the corresponding values of M and θ in the four quadrants we obtain from $\Sigma M/\Sigma\theta$ mean values of M/θ at various angles of displacement. These results are given in table 3 and are well represented by the relationship

$$M/\theta=124{\cdot}932+3{\cdot}665\times 10^{-4}\,\theta^{2}-3{\cdot}05\times 10^{-7}\,\theta^{4}$$

as will be seen from the fifth column, which gives the values calculated from this expression. The differences between the observed and calculated values of M/θ are given in the last column.

This setting of the coils with a maximum value for M/θ of 125·042 nominal μH. per degree at 24°·50 appears ideal for use with a commutator having sector contacts of semi-angle between 23° and 26°.. Accordingly the commutator M, figure 2, with smaller sector of angle tapering from 44° to 51° was adjusted on the shaft.

Determination of the constant K' for large angle of contact. The larger commutator was adjusted systematically and firmly fixed in position. The constant K' of equa-

Table 3. Values of M/θ at various angles θ. $x/a = 0{\cdot}212$, $M_{\text{max.}} = 7249{\cdot}37$

M (nominal (μH.)	(θ seconds of arc)	θ (degrees)	M/θ observed (nominal μH./degree)	M/θ calculated (nominal μH./degree)	Difference
3261·2	93,961	6·53	124·949	124·947	+0·002
4842·7	139,511	9·69	124·963	124·964	−0·001
7247·3	208,722	14·49	125·000	124·996	+0·004
9690·8	279,048	19·38	125·021	125·027	−0·006
10957·8	315,489	21·91	125·038	125·038	Zero
12255·2	352,824	24·50	125·045	125·042	+0·003
13625·1	392,298	27·24	125·033	125·036	−0·003
14927·4	429,860	29·85	125·014	125·016	−0·002
16834·9	485,007	33·68	124·958	124·955	+0·003

tion (5), equivalent to $(M'/\theta' + M''/\theta'')/2$, was found for it directly *in situ* by taking readings of M and θ at the sector edges over all four quadrants with the brushes (*a*) just on and (*b*) just off. The data are recorded in table 4.

Table 4. Direct determination of K' for contact of 47°

$x/a = 0{\cdot}212$, $M_{\text{max.}} = 7249{\cdot}37$

Quadrant	Contact	M (nominal μH.)	M_θ (nominal μH.)	θ (seconds of arc)
1	On	2945·7	2890·0	84,579
1	Off	2968·1	2910·7	85,222
2	On	2924·6	2885·9	84,451
2	Off	2927·2	2888·1	84,520
3	On	2880·8	2827·8	82,653
3	Off	2888·3	2835·0	82,876
4	On	2937·2	2898·4	84,840
4	Off	2953·0	2912·5	85,278

The mean value of K' for the on and off positions, which owing to the closeness of the ratios involved is equivalent to $\Sigma M/\Sigma\theta$ for all readings, is $125{\cdot}040_5$ nominal μH. degree or, when corrected by the factor 1·00035 to convert to true microhenries, 125·084 μH. degree for a semi-angle of contact of 23°·4.

Spin experiments with an angle of contact of 47°. A series of spin experiments was performed in the manner already described, and the balancing value of R in c.g.s. units was given by

$$R = 36 \times 125{\cdot}084 \times 10^4 \times n.$$

Table 5 shows the values of R so obtained and the corresponding potentiometer checks. The synchronized motor was used in the first three experiments.

Table 5. Absolute measurement of resistance. $K' = 125{\cdot}084$ μH./degree

Frequency of fork (c./sec.)	Sectors on disc	n (rev./sec.)	$R \times 10^{-9}$ (c.g.s. units)	Resistance by comparison (Ω.)
256	30	8·533	0·38426	0·38398
320	30	10·667	0·48032	0·48018
384	30	12·800	0·57639	0·57614
256	24	10·667	0·48032	0·48008
320	24	13·333	0·60040	0·59982
256	36	7·111	0·32022	0·32047
256	36	14·222	0·64043	0·64036
512	36	14·222	0·64043	0·64026
320	36	8·889	0·40027	0·40002
384	36	10·667	0·48032	0·48033
512	40	12·800	0·57639	0·57627

§ 5. CONCLUDING REMARKS

As this method of measuring resistance absolutely is based on the constancy of the ratio M/θ in the neighbourhood of an angle θ which determines the approximate semi-angle of the contacts employed, the greater part of our time has been given to an investigation of the relationship between M and θ and to devising means of measuring the ratio accurately. Once the laws of inductance made use of have been established, a few readings suffice to determine the constant K for any commutator in position and to set its value between close limits. We favour the use of contacts of some 50° of arc owing to the great accuracy with which K can then be measured.

In view of the high sensitivity and the rapidity and ease with which balance can be obtained, we propose to undertake further work with coils of rather greater diameter (allowing more space between the fixed pair and the rotator) so wound that resistances of the order of an ohm may be conveniently measured. Further investigation is also being undertaken of the limits of accuracy of the sine law for the angle solenoid. It is proposed for the purpose of measuring θ to use two additional twin coils always in the Helmholtz position and possessing a maximum mutual inductance of some 10 mH. with the angle coil. The essential length and time measurements will then be dependent upon the value of a convenient standard inductance and upon a standard tuning-fork frequency.

§ 6. ACKNOWLEDGMENTS

We express our gratitude to Prof. P. M. S. Blackett, M.A., F.R.S. for providing us with facilities for carrying out this research and for the encouragement he has given us throughout the investigation. We are indebted to Mr S. Baker for the design and construction of the fork-controlled motor-synchronizing unit and to Mr H. G. Bell for valuable help in the construction of the apparatus.

DISCUSSION

Dr D. Owen. The method described is not really related to that of the B.A. revolving coil, as appears to be implied in the introduction to the paper. Its affinity is rather with the Lorentz method, as is clearly indicated by the formula $R=Mn$ which applies to both. The idea of using a momentary contact at the instant of maximum induced e.m.f. of the moving conductor was proposed by Lippmann, who used a coil rotating in the uniform field within a solenoid carrying the same current as that passing through the resistance to be measured. The advantage of the present method lies in the application of the investigation previously made by one of the authors of a type of variable mutual inductance in which, over a wide range of angular movement, the mutual inductance is very closely a linear function of the angle. This at once puts the determination of resistance on an altogether higher plane of accuracy. Compared with the Lorentz revolving-disc method, it is now possible to use a multilayered coil, and consequently the scale of size of the whole apparatus, or the speed of rotation of the coil, may be greatly reduced. These advantages may well engage the careful consideration of those concerned with future work on the determination of the ohm at the various national laboratories.

Authors' reply. The method here described resembles the B.A., the Carey Foster and the Lippmann methods in that an alternating e.m.f. is generated in the revolving coil. It differs from them in that the e.m.f. is not sinusoidal and is very uniform over the contacts particularly at the contact edges. In Lorentz's method the magnetic lines of force are cut throughout a revolution at a constant rate and the e.m.f. is unvarying.

We thank Dr Owen for drawing our attention to Lippmann's method, the formula for which may be written $R=Mn$ if the contact is momentary. Our remarks in the paper on the correction for arc of contact in the Carey Foster method are equally applicable to the Lippmann method.

535.62

A NEW PRECISION COLORIMETER

By JOHN J. MANLEY, M.A., D.Sc. Oxon.,
Fellow of Magdalen College, Oxford

Received June 10, 1934. *Read in title November* 2, 1934.

ABSTRACT. In this paper is given a description of a colorimeter (or tintometer) in which use is made of a simple optical system for juxtaposing the two beams of light. Also, novel methods are introduced for equalizing the tints of the two columns of liquid and for determining the difference between their lengths.

---

§ 1. INTRODUCTION

Of late I have had occasion to carry out many determinations of minute quantities of iron, and for that purpose the precision colorimeter which is the subject of this paper was designed and constructed. The chief features to which attention may be drawn are (i) the simplicity and consequent inexpensiveness of the optical system; (ii) the plan used for varying the length of the column of the standard solution; and (iii) the differential method employed for ascertaining the relative values of the two solutions contained in the tubes *A* and *B*.

§ 2. CONSTRUCTION OF APPARATUS

Figures 1, 2 and 3 are largely self-explanatory. It may be seen that the several glass and other parts of the instrument are, as usual, mounted upon a vertical board *C*. The tubes *A* and *B*, graduated in centimetres, are supported by a shelf *D* having suitable apertures for the transmission of light reflected from the mirror *M**. Variations in the inclination of the mirror are producible by means of the screw *S*. On emerging from the tubes *A*, *B*, the two beams of light enter and pass through the V-shaped block of glass *G*, *G*, and are thus brought close together. Each limb of the vee is 18 mm. thick and the faces are polished and parallel. The two plates, united with Canada balsam, meet and form as shown in the figure an angle of 78°. After traversing the prismatic block the beam of light is, by means of a diaphragm *F*, *F* having an opening 4 mm. long and 3 mm. wide, so limited that the central portions only pass upwards into the eye-piece *E*. The level of the standard solution in *B* can be varied and adjusted by a glass plunger *P* contained within a cistern *T*. The plunger, of uniform diameter, is suspended from a pin *N* by a platinum ribbon *O*

---

* The tubes *A* and *B* are not enclosed. Theory and experiment alike show that the use of a cover to exclude light other than that reflected by the mirror is wholly unnecessary. A decided advantage, however, accrues from the screening of the prismatic block for which an aluminium cover is used.

0·1 mm. wide and 0·08 mm. thick; and it is so loaded with mercury that even when it is immersed to the fullest extent the ribbon still remains taut. For keeping the plunger central three solid glass pimples *a*, *b*, *c* were formed and then attached by fusion as shown. In the absence of these the plunger makes contact with the walls of the cistern, thus causing inconvenience and inaccuracy. Two stops *d*, *e* impose a limit to the degree of rotation of the disc *X* carrying the index *I*. A ring *L* having upon its outer rim a millimetre scale as shown can be rotated independently of *X* and then secured in any one of a number of positions by a pin *q* fixed to a steel spring *V*. For convenience, the diameter of the cylinder *Y* and that of the graduated ring *L* are so proportioned that when *X* is rotated and *P* raised or lowered until the level of the solution in *B* has been changed by ∓ 10 mm., the reading as shown by *I* is altered by ± 20 scale divisions. Under these circumstances each scale division (1 mm.) corresponds to a variation in the level of the liquid in *B* equal to ∓ 0·5 mm. Smaller quantities are a matter of estimation. The plunger is retained by friction in any position to which it may be brought, and the friction, which is but small, is produced by an appropriate adjustment of the nut *U*.

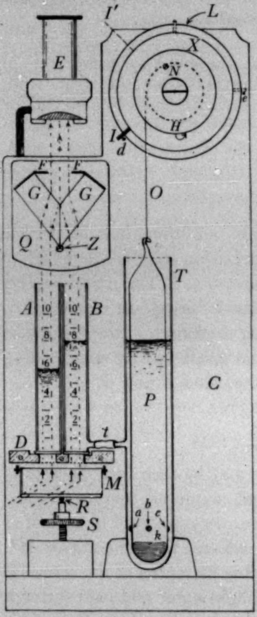


Figure 1.

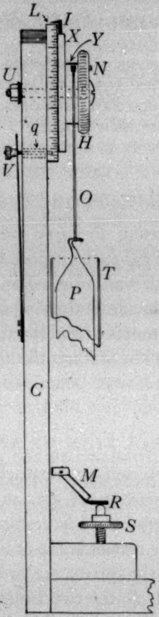


Figure 2.

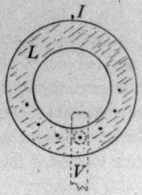


Figure 3.

In figure 3 is shown the posterior surface of the ring *L* with its series of equally spaced cavities* for the reception of the pin *q*. The distance between any two cavities is such that it is equivalent to a difference of 0·2 unit (cm.) of the scale upon either glass tube, and to 4 scale divisions of the ring *L*.

The plate *Q* is pivoted upon the pin *Z*, and thus any required adjustment of the prismatic block about a horizontal axis can be readily effected; the plate can then be clamped by means of a nut (not shown) at the back of the support.

For attaching the parallel-plane plates to the tubes *A* and *B*, use was made of a cement sold under the name of "Durofix." This cement, which is not affected by water, has proved highly satisfactory.

§ 3. EXPERIMENTAL PROCEDURE

The colorimeter is used in the following way. First, the coloured solution under examination is introduced into *A* and the level of the liquid adjusted so that it coincides with an appropriate division, say 3, 6 or 8. The most convenient length for the column obviously depends upon the intensity or depth of the colour. Let us suppose the length to be 5 cm. Secondly, the plunger *P* is moved by turning *H* until the index *I* is in the position indicated by *I′*. Thirdly, *V* is now pulled back so as to withdraw the attached pin *q* from its cavity: then the graduated ring *L* is rotated until the zero of its mm. scale is in close proximity to the index at *I′*. On releasing *V* and slowly turning *L* to the right or left *q* springs into an adjacent cavity and thus for the time being maintains the ring in a fixed position. The index is then set at zero. Fourthly, the standard solution is poured into the cistern *T* until the level of the liquid in *B* coincides with the graduation 4 (not with 5 as in the case of *A*). By adopting this plan, the level of the liquid in *B* can by means of the plunger be varied within the limits $5+2$ and $5-2$ cm. Lastly, by means of the milled head *H* the plunger is alternately raised and lowered until finally the two patches of colour as seen through the eye-piece are precisely matched. The index reading is then noted, and as this at once gives the difference in the levels of the two columns the required value may then be calculated.

It will be observed that differences measured in the way just described assume values which are positive when the plunger is lowered and negative when it is raised. Accordingly all graduations of the scale on the left of zero as seen from the front are marked +, whilst those on the right are marked −.

It may be desirable to point out that if some level other than 5 be more convenient for the measurement in hand the procedure remains unchanged. For example, let the length of the column in *A* be increased to 7. This column is 2 cm. longer than the former, and as each 1 cm. is equivalent to 20 divisions of the graduated ring, the ring must be rotated through 2×20 to the right and then, as in the first case, secured by the pin *q*. Next, the index *I* having been brought to the zero, an additional quantity of the standard solution is introduced into the cistern *T* and the level of the liquid in *B* is adjusted so as to coincide with the graduation marked

---

* Actually the cavities are five times more numerous than those represented in the figure.

6. That is with 7 − 1 instead of with 5 − 1 as in our first illustration. Finally the two colours are matched as already described and their relative values calculated.

When the instrument is to be used for liquids which are affected by rubber the side-tubes at *t* may be united by fusion.

§ 4. ACCURACY ATTAINABLE

To those experienced in the practice of colorimetry, it will be quite obvious that the degree of accuracy attainable with the instrument just described must be, and is, of the same order as that ensured by the use of other precision colorimeters. And here we remark that the chief difficulty associated with the use of colorimeters arises not so much from the form of any given instrument as from a lack, on the part of the observer, of a necessary sensitivity for apprehending differences in colour-density. Some observers are able to match particular colours with ease but experience considerable difficulty when dealing with others. In this respect my own observations are uncertain and irregular in the case of yellow, pink or red solutions, but are reasonably uniform for those which are blue. And so for the research which, as we have already stated, involved the determination of minute quantities of iron, choice was made of Prussian blue rather than of the red compound iron-potassium thiocyanate. Curiously enough a friend engaged in a similar research prefers the latter to the former compound.

For ascertaining the mean error attending the use of the new instrument, the following experiments were made. First an acid solution of iron in the ferric state was prepared, the concentration of which was 56 mg. of iron per 100 cm.^{3} Of this solution 1 cm.^{3} was transferred to a 100 cm.^{3} measuring-flask and freely diluted. Then the required potassium ferrocyanide was added, and the whole was well mixed and finally made up to the standard volume. Newly prepared, such a dilute solution remains quite unclouded for some days.

Next, both tubes of the apparatus were charged with standard solution and the necessary settings and adjustments carried out in the way already described. For the first experiment the column of liquid in *A* was 40 mm. long. Finally, column *A* was repeatedly measured in terms of column *B* with the results set forth in table 1.

Table 1

Length of column (mm.)			
Observed		True	Differences (mm.)
Plunger raised	39·5	40·0	−0·5
	40·5	40·0	+0·5
Plunger lowered	39·5	40·0	−0·5
	41·0	40·0	+1·0
Mean = $40{\cdot}1_3$			Mean = $+0{\cdot}1_3$

The mean difference $+0{\cdot}1_3$ is equivalent to +0·065 mm. (*vide supra*). Hence the measured weight of iron differs from that known to be present by $\frac{0{\cdot}56 - 0{\cdot}56 \times 40}{40{\cdot}065}$ or $0{\cdot}56 - 0{\cdot}559_8$ mg.

In continuation of the series of tests, *A* and *B* were again charged with the solution containing 0·56 mg. of iron per 100 cm.^{3} The procedure detailed above was again followed, and other 3 columns of the standard solution were measured with the results set forth in table 2.

Table 2

Length of column in tube *A* (mm.)	Mean measured length of column in tube *B* (mm.)	Iron content		
		True (mg.)	Observed (mg.)	Differences (per cent)
60	59·62	0·56	0·5564$_5$	−0·64
80	79·88	0·56	0·5591$_7$	−0·14
100	100·01	0·56	0·5600$_6$	+0·04
			Means = 0·5586	−0·25

Finally, with the object of rendering these tests more complete, several solutions having concentrations less than that of the standard liquid were prepared and measured with the results shown in table 3.

Table 3

Solution no.	Length of column in *A* (mm.)	Iron content		
		True (mg.)	Found (mg.)	Difference (per cent)
1	90	0·187	0·1865	−0·27
2	90	0·250	0·2485	−0·60
3	60	0·280	0·2770	−1·08
4	60	0·373	0·3720	−0·27
			Mean =	−0·56

From tables 2 and 3 it may be seen that the final mean error is of a negative character and that therefore, in general, the weight of the iron was underestimated. But given another observer, having a personal equation differing from my own, the corresponding mean error might well be dissimilar not only in magnitude but also in sign.

§ 5. EFFECTS OF EYE-FATIGUE

In conclusion, attention may be drawn to the way in which the effects of eye-fatigue resulting from an indispensable critical observation of the colour-patches were met and overcome. Experiment showed that for a series of readings the greatest uniformity in apparent values was obtained when the two columns were balanced quickly, or by first approximating to the true value and then resting the eye before the final adjustment was made. Another and most helpful plan was that in which the main adjustment was carried out with the aid of one eye and the final and more precise setting accomplished with the aid of the other and unfatigued eye. The apparent increase in the brightness of the field, as judged by the second eye, is usually very considerable, and so any residual difference in the depth of colour of the two columns is the more readily detected and balanced. I am not aware that these important matters of detail have hitherto received the notice they seem to merit.

537·565:546.293

THE ENERGY OF AGITATION OF POSITIVE IONS IN ARGON

By F. LLEWELLYN JONES, M.A., D.Phil.,
University College of Swansea

Communicated by Prof. E. J. Evans, July 2, 1934. *Read in title, November* 16, 1934.

ABSTRACT. The energy of agitation of positive argon ions was investigated for a range of electric forces Z from 5 to 50 V./cm. and for a range of pressures p from 3 to 0·24 mm. by means of the Townsend method of measuring the lateral diffusion of a beam of ions moving in a uniform electric field.

In pure specimens of gas the energy of agitation of the ions was the same as that of the atoms of the gas, and independent of the pressure, when the ratio Z/p was less than about 20. For higher values of Z/p the mean energy was increased by the field; the values obtained are compared with those calculated on the assumption that the ions behave as perfectly elastic spheres which attract the atoms of the gas with a force varying as the inverse fifth power of the distance.

The motion of the ions was found to be very sensitive to the presence of impurities, and the general effect of the impurities at pressures greater than 0·4 mm. and when Z/p was less than 80 was to reduce the mean energy of agitation of the ions to the value possessed by the atoms of the gas. This is probably due to the formation of ionic clusters.

---

§ 1. INTRODUCTION

When a stream of positive ions moves through a gas at a pressure of p mm. in a uniform electric field Z their mean energy of agitation is not very different from that of the molecules of the gas when the ratio Z/p is less than about 10*. When Z/p is increased, there are various processes which determine whether the mean energy of the ions also increases. The influence of these processes has been fully discussed by Tyndall and Powell† in their investigation of the mobilities of positive ions in pure gases. If the gas contains impurities, molecules of the impurity may collect around the monatomic ion to form a cluster of many times atomic magnitude. This would especially be the case at high gas pressures and if the impurity molecules have a permanent dipole. Owing to their large mass the mean energy of agitation of these clusters would then be the same as that of the atoms of the gas, even for higher values of the ratio Z/p of the order of 100. However, as the pressure is reduced the influence of impurities decreases rapidly. The chance of a cluster being formed will decrease, and, as the electric intensity is increased the clusters will be broken up by collisions with atoms, and the ion will become monatomic. Again, the process of electron-transfer from a neutral atom to the positive ion with which it collides is considered by some physicists to modify

* Z expressed in V./cm.

† A. M. Tyndall and C. F. Powell, *Proc. roy. Soc.* A, **129**, 162 (1930); A, **134**, 125 (1931).

the energy of agitation of the ions, owing to the fact that the charge may not remain associated with the same ion as it moves through the gas*. This process is the more important in the case when the ions are those of the gas through which they move. It is suggested† that electron-transfer introduces short-range interaction forces between the ions and atoms of the gas, so that the total force between a positive ion and gas atoms must contain these exchange forces in addition to those due to the polarization of the neutral atoms by the charge on the ions. The mobility of the ions, and hence their energy of agitation, depends on these forces. Hassé and Cook‡ have shown, however, that if the collisions between the ions and atoms resemble those between elastic spheres the transfer of charge makes no difference to the mobility of the ions, but in general, when the ion exerts forces on the atoms, the mobility is reduced by the process.

§ 2. DESCRIPTION OF APPARATUS

In order to investigate the variation of the mean energy of agitation of positive ions with the ratio Z/p the following experiments were performed in argon. The best method for the determination of the energy of agitation of ions in gases in uniform electric fields is that due to Townsend§, which consists in measuring the lateral diffusion of a narrow beam of ions moving in a uniform electric field between two parallel plates.

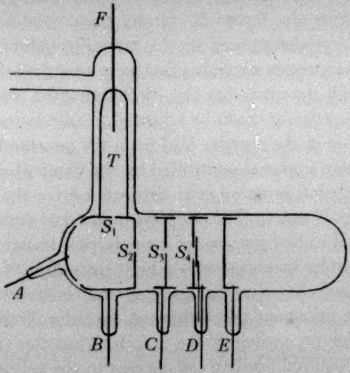


Figure 1.

The apparatus employed is shown diagrammatically in figure 1. The design of the electrodes C, D and E was similar to that used in an apparatus for measuring the ionization produced by positive ions described previously*. They were thin discs 3·4 cm. in diameter, fitted with cylindrical rings, and mounted 1·8 cm. apart on three pyrex rods. The slit S_3 in the electrode C was 1·25 mm. wide and 5 mm. long, and the slit S_4 in D was 2·5 mm. wide and 6·5 mm. long. A shutter forming part of the electrode D could be placed over the slit or withdrawn to the side to leave it open, and was operated magnetically.

---

* F. M. Penning and C. Veenemans, *Z. Physik.* **62**, 746 (1930).

† J. E. Lennard-Jones, *Proc. phys. Soc.* **43**, 461 (1931).

‡ H. R. Hassé and W. R. Cook, *Phil. Mag.* **12**, 554 (1931).

§ J. S. Townsend, *Proc. roy. Soc.* A, **81**, 464 (1908) and *Motions of Electrons in Gases*, Clarendon Press, pp. 7–9.

The positive ions which pass through these slits were generated in a side tube by means of a direct-current discharge between the molybdenum wire anode F and the electrode B. A fraction of the ions which passed through the slit S_1 were repelled by the electrode A and passed through the slit S_2 and finally entered the space between C and D. Slits S_2, S_3 and S_4 were in the same straight line. Ions were prevented from reaching the electrode C except through the slit S_2 by confining the discharge within a glass tube T, 6 cm. long and 1·5 cm. in diameter, which fitted tightly into a collar fixed to the cathode. Also, no ultraviolet light from the discharge in T could pass through the slits S_3 and S_4. The electrodes were made of copper 0·5 mm. thick and were insulated by pyrex bars. The glass envelope was also of pyrex and the connections to the electrodes were brought out through molybdenum seals.

The electrode C was maintained at earth potential, and the potential difference between A and B was about 10 volts. The electric force Z between the electrodes E and D was the same as that between D and C, which was approximately the same as that between the electrodes C and B. Under these conditions the positive ions attained the mean energy determined by the ratio Z/p before entering the slit S_3. The potentials of the electrodes were obtained from a battery of small accumulators.

The source of high potential for the discharge tube was a 600-volt direct-current generator. The current could be adjusted to any desired value by means of a diode valve interposed in the positive lead from the generator. Thus the intensity of the current of positive ions was controlled by the filament rheostat of the diode, and their mean energy of agitation was determined by the potential-difference between the electrodes B and C. The simplest method of cutting off the stream of positive ions was found to be by means of a single-pole switch placed in the anode lead to the discharge tube in series with a high resistance of the order of 1 MΩ. In this way the difference between the sparking and the maintenance potentials of the discharge was set up across the resistance, and the discharge current always attained the same value on closing of the switch. Thus the apparatus was readily adjusted to give any required positive ion current to the collectors.

§ 3. PURIFICATION OF GAS

The experiments were made with argon which was obtained commercially pure from the General Electric Company, Ltd. Further purification was effected by circulating the gas over red-hot calcium, in the manner devised by MacCallum and Klatzow†. The calcium was contained in a silica tube which could be raised to red heat electrically. This tube was connected to the apparatus by means of a ground-glass joint, vacuum grease being used on only the outer half of the seal. The diffusion of mercury vapour from the Mcleod gauge and vapour pump into the pure gas was prevented by means of a liquid air-trap in the usual manner.

---

* J. S. Townsend and F. Llewellyn Jones, *Phil. Mag.* **15**, 282 (1933).

† S. P. MacCallum and L. Klatzow, *Phil. Mag.* **15**, 829 (1933).

The pyrex apparatus was heated for several hours to expel gases from the electrodes and from the surface of the glass, and a discharge was also maintained in the side tube. The argon was kept in contact with the red-hot calcium for some time before entering the apparatus. The gas was then withdrawn by absorbing it in a charcoal trap cooled by liquid air, again passing over the heated calcium. This operation was repeated many times and the argon was stored over the calcium while it cooled to room-temperature before being finally admitted to the apparatus for the measurement of the currents. The circulation of the gas over the hot calcium was also repeated during the course of the experiments.

§ 4. EXPERIMENTAL PROCEDURE

The discharge current between the electrodes A and B had a value between 60 and 120 μA., depending on the pressure of the gas, while the currents of positive ions passing through the slit S_3 were of the order of 10^{-11} A. Larger currents are undesirable in order that the space charge may not distort the electric field between the electrodes. Under these conditions the currents were measured by observing the charges received by the electrodes D and E when a discharge was maintained for 10 seconds in the side tube. The charges were measured by an electrostatic induction balance consisting of a quadrant electrometer, screened air condensers, and a potentiometer, in order that the potential of the insulated electrode may not change while the electrode was receiving a current of positive ions.

At some pressures it was found that a visible stream of positive ions passed through S_1 into the space between A and B. Some of the ultra-violet light emitted passed through the slit in the electrode C and caused the emission of electrons from D and E. In order to estimate the currents due to the positive ions it is necessary to correct the observations by deducting from the charges received by the discs D and E the small charges received when the electrodes B and C were at the same potential, for then the positive-ion current passing through the slit S_3 was negligible. These corrections were of the order of 5 per cent of the positive ion current at the higher pressures but rose to 20 per cent at the lowest pressure with the higher forces.

§ 5. RESULTS OBTAINED

As the ions move from C to E under the electric force Z the beam diverges so that some fall on the electrode D, and the rest, when the slit S_4 is open, pass through to be collected on E. When the shutter closes the slit, the whole beam of ions is collected on electrode D. The experimental part of the investigation consisted in measuring the currents i_1 to the electrode E, the current i_2 to the electrode D when the slit was open, and the current i_3 to D when it was shut. If all the ions which pass through the slit are collected on E, then $i_3=i_1+i_2$; and this was found to be the case in all the measurements. The ratio i_1/i_3 was determined at various pressures of gas from 3 mm. to 0·24 mm. for a range of electric forces from 10 to 50 V./cm.

The results of the experiments with the purest specimens of argon are shown by the curves in figure 2, in which the ordinates represent R, the ratio i_1/i_3, while the abscissae represent the electric intensity.

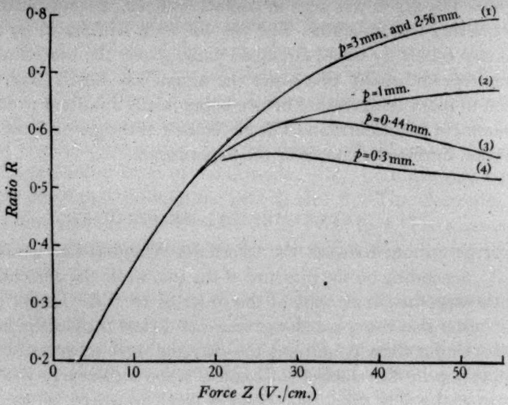


Figure 2.

Curve (1), obtained at a pressure of 3 mm., may be taken to correspond to the case when the energy of the ions is the same as that of the atoms of the gas, the ratio Z/p being less than 17. When the pressure was reduced to 1 mm. curve (2) was obtained, and for values of Z/p less than 20 it is identical with curve (1). This shows that values of Z/p less than about 20 have little effect in increasing the energy of agitation of the positive ions. However, for higher values of Z the divergence of the ions increased, indicating an increase in the mean energy. Curve (4), obtained at a pressure of 0·3 mm., shows that the energy of agitation was increased still further, and the ratio R at about 0·5 was independent of higher values of the electric force.

However, during the course of the experiments it was found that the presence of minute quantities of impurities had a large effect on the energy of the positive ions. Thus at a pressure of 1 mm. a specimen of gas was allowed to remain in the apparatus and was not circulated over the calcium during the measurement of the currents. The curve thus obtained practically coincided with curve (1), showing that the energy of agitation of the ions was the same as that of the atoms of the gas for values of Z/p up to 60. Highly purified argon has been prepared by MacCallum and Klatzow*, who found that the presence of impurities in minute quantities has a large effect on the sparking-potential and the electrical properties of the gas.

---

* *Loc. cit.*

Experiments were also performed with positive ions in specimens of argon which were known to be impure. Purified argon at a pressure of 1·52 mm. was admitted to the apparatus and frequent discharges were produced in the side tube, currents of 200 μA. being used in order to cause the emission of impurities. The gas was not circulated over the heated calcium for a period of many weeks. Measurements of the ratio R were made for a range of pressures between 1·52 mm. and 0·24 mm. The results are given in the curves of figure 3.

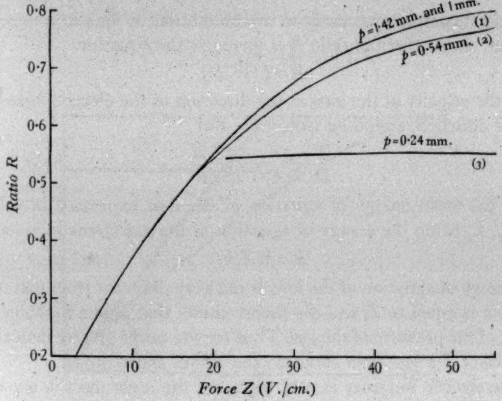


Figure 3.

For pressures of 1·5, 1·0 and 0·54 mm. the curves are all very close together, all roughly corresponding to the case when the energy of the ions was equal to that of the gas atoms. Thus the presence of the impurities appears to prevent the energy of agitation of the ions from becoming greater than that of the atoms of the gas for pressures greater than about 0·5 mm. with electric forces less than 60 V./cm. However, when the pressure was reduced to 0·24 mm. the energy of the ions was increased by the electric field just as in the case of the purer specimens of gas. A reasonable explanation of this effect is the increase in the mass of the ions due to the clustering of molecules of the impurity around them; and the experiments show that these clusters may exist even when the ratio Z/p is about 80. It would thus appear that impurities have much less effect at the low pressures of the order of 0·25 mm. in reducing the energy of agitation of the ions. This is doubtless due to the smaller chance of the formation of clusters, and the strong electric field breaking them up.

By comparing the curves for a pressure of 1 mm. obtained with the pure and impure specimens of argon it is seen that the effects produced by the impurities appear not to depend very greatly on the quantity of the impurity present. Thus the curve for argon containing only traces of impurities—approximately curve (1), figure 2—is practically the same as the curve obtained with the very impure specimens, curve (2), figure 3. The effect due to the first trace of impurity is very great. The high sensitivity to impurities of the motion of the positive ions in argon makes it probable that the values of the ratio shown by curves (2), (3), (4) in figure 2 are too high, that is that the electric field increases the mean energy of agitation of the ions more than is indicated in these experiments.

§ 6. INTERPRETATION OF RESULTS

Townsend's* well-known analysis of the motion of ions in a uniform electric field Z V./cm. shows that the ratio R is given by the equation

$$R=f(W/D) \qquad \text{......(1),}$$

W, D where W is the velocity of the ions in the direction of the electric force and D their coefficient of diffusion, supposed isotropic. But

$$\frac{W}{D}=\frac{3Z}{2E_1}=\frac{3}{2E_2}\cdot\frac{Z}{k} \qquad \text{......(2),}$$

E_1 where E_1 is the mean energy of agitation of the ions expressed in volts which is
k, E_2 equal to kE_2, E_2 being the energy of agitation of the gas atoms in volts. Thus

$$R=f(Z/k) \qquad \text{......(3).}$$

When the energy of agitation of the ions is not very different from that of the atoms of the gas, Z/k is equal to Z, and the theory shows that R is a function of Z and is independent of the pressure of the gas. Thus for pressures greater than about 1 mm. and for values of Z/p less than about 20 the curves are identical.

When the electric intensity is very high and the ionic mass is equal to that of an atom of the gas, the drift velocity W of the ions may become comparable with
u their velocity of agitation u. Consequently the equations of diffusion would be invalidated and equation (3) would require modification. In this case the velocity of the ions is principally in the direction of the electric force, and their energy is due to this motion and not to their motion of agitation. Hence in these experiments the ratio Z/p was limited to values less than 150. For large values of Z the free paths of the ions are bent appreciably in the direction of the field, and the lateral diffusion of the ions is reduced. Thus there is an apparent reduction in the coefficient of lateral diffusion D, and since the energy of agitation of the ions is obtained from this coefficient by means of the equations (1) and (2), it follows that the values of k calculated from equation (3) are too low when Z/p is very large.

An approximately correct modification of the equations may be obtained simply in the case when the collisions between the ions and atoms resemble those between elastic spheres.

In the absence of the electric force let OA, figure 4, represent the free path of
τ an ion after a collision at O. The mean interval, τ, between successive collisions is
$\lambda, \bar{u}$ equal to $\lambda/\bar{u}$, where λ is the mean free path and $\bar{u}$ the mean velocity of agitation. When a strong field Z is acting, in the absence of a collision during an interval τ
s the ion traverses the parabolic path OCB of length s.

---

* *Loc. cit.*

Since the mean free path is practically unaltered, the ion now collides at C, where $OC=\lambda$. The lateral diffusion of the ion is thus reduced from OY to OY'. The fractional reduction f is OY'/OY, which, though less than OC/OB, is practically equal to it, i.e. to λ/s. If $\bar{u}_1$ is the mean velocity with which the ion traverses the parabolic path s, then $s=\bar{u}_1\tau$; hence

f

$\bar{u}_1$

$$\lambda/s=\bar{u}/\bar{u}_1.$$

Now

$$\bar{u}_1=[\bar{u}^2\sin^2\theta+(\bar{u}\cos\theta+\tfrac{1}{2}\gamma\tau)^2]^{\frac{1}{2}},$$

where $\gamma=eZ/m$, e being the ionic charge and m the mass of the ion. Thus

γ, e, m

$$f=[1+\gamma\tau\cos\theta/\bar{u}+(\tfrac{1}{2}\gamma\tau/\bar{u})^2]^{-\frac{1}{2}}.$$

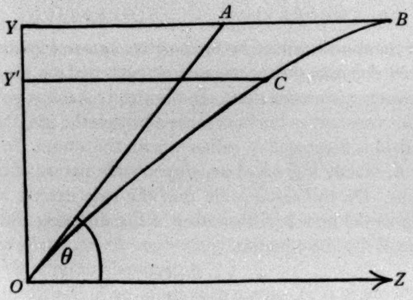


Figure 4.

When the velocity of the ions in the direction of the electric force is large, no great error is made in replacing the mean value of $\cos\theta$ by unity, so that $\bar{f}$ is given by

$$\bar{f}=\frac{1}{1+\frac{1}{2}\gamma\tau/\bar{u}}=1\bigg/\left(1+\frac{Z}{p}\cdot\frac{\lambda_1}{4kE_2}\right),$$

where λ_1 is the mean free path of the ions at a pressure of 1 mm. and is computed below* as a function of k, equation (5). The coefficient D_1 of lateral diffusion of the ion is proportional to the mean free path of the ions in a direction perpendicular to the electric force and is thus given by

λ_1

D_1

$$D_1=D/(1+C.Z/p) \qquad \text{......(4)},$$

where C is a function of k, and is $4{\cdot}8\times10^{-3}$ when k is 5. Thus the correction necessary is less than about 30 per cent when Z/p is 60.

C

It may be seen† that when the coefficients of diffusion along and perpendicular to the electric force are not the same that equation (1) becomes

$$R=f(W/D_1).$$

---

* See p. 82.

† See for example V. A. Bailey, *Phil. Mag.* **9**, 560 (1930).

Hence the values of k obtained from (3) should be multiplied by $(1+C.Z/p)$ to obtain the energy of the ions. These corrected values of k are given in the table for the different values of the ratio Z/p used in the experiments.

Table 1

k	1	1	1·5	2	2·7	3·8	5
Z/p observed	0	18	27	45	75	100	150
Z/p calculated from equation (10)	0	14	30	43	60	80	100

§ 7. THEORETICAL CONSIDERATIONS

E_1

The energy of agitation acquired by the positive ions in a gas under the influence of the electric field depends on the mean free path and on the fractional loss of energy in an encounter with a gas atom. In the steady state, when the mean energy of agitation remains constant as the ions move through the gas, the energy acquired from the electric field is dissipated in collision with the atoms. In order to calculate this mean energy E_1 exactly it is necessary to know the nature of the forces between the ions and atoms. On the assumption that the ions exert a repulsive force on the atoms varying as the inverse fifth power of the distance, Pidduck* has shown that k is related to W by the equation

$$k-1=W^2/\Omega^2,$$

Ω

where Ω is the velocity of agitation of the gas atoms. However, it is interesting to estimate the energy of the ions on the assumption that their collisions with gas atoms resemble those between elastic spheres, and that in addition they attract the atoms with the force due to the polarization of the gas atoms by the ions. In this case the attractive force between the ion and an atom due to the charge induced on the atom has been given by Langevin† in the form

$$F(r)=\frac{(K-1)\,e^2}{2\pi N r^5}=\frac{B}{r^5},$$

r, K
N
λ, λ'

where r is the distance between the centres of the atom and the ion and K the dielectric constant of the gas containing N atoms per unit volume. The rate of loss of energy of ions in a gas under these conditions has been investigated by J. J. Thomson‡ who showed that the increase in the energy-loss due to the charge on the ions may be regarded as being equivalent to a reduction in their effective mean free path. Let λ be the mean free path of the ions and λ' the mean free path of the atoms with the same energy. Then if the attractive force is strong

$$\lambda=\frac{\lambda'}{4{\cdot}4\,(\omega/E_r)^{\frac{1}{2}}} \qquad \text{......(5)},$$

---

* F. B. Pidduck, *Proc. roy. Soc.* A, **88**, 296 (1913).
† P. Langevin, *Ann. Chim.* (*Phys.*), **8**, 245 (1905).
‡ Sir J. J. Thomson, *Phil. Mag.* **47**, 337 (1924).

where ω is the potential energy of an atom in contact with an ion, and E_r is the relative kinetic energy of the ion. Thus

ω, E_r

$$\frac{\omega}{E_r}=\frac{B}{4\sigma_{12}{}^2E_r},$$

where σ_{12} is the sum of the radii of an atom and an ion.

σ_{12}

In deducing this relation it is assumed that the force on the atoms of the gas due to the charge on the ions is strong enough to bend the paths of the ions or atoms so that collisions between them are directly along the line of centres, and also that exchanges of momentum occur between the ions and the atoms when they do not collide. In general the rate of loss of energy by the ions is less than that deduced from these assumptions, but the error introduced in adopting equation (5) becomes smaller the more polarizable are the atoms of the gas.

When the positive ions are not in thermal equilibrium with the gas atoms, their mobility μ may be found from their coefficient of diffusion D by means of Langevin's relation given in equation (2). If the velocities of the ions are distributed according to Maxwell's law, D is given by the equation*

μ, D

$$D=\frac{4}{(3\pi)^{\frac{3}{2}}N\sigma_{12}{}^2}\left(\frac{E_1}{m_1}+\frac{E_2}{m_2}\right)^{\frac{1}{2}} \qquad \text{......(6)},$$

E_1 and E_2 being the mean energies of the ions and atoms measured in ergs. When the ions are those of the gas through which they move, then $m_1=m_2=m$ the atomic mass and $\sigma_{12}=\sigma$ the atomic diameter.

E_1, E_2

σ

The mean free path λ' of an atom with energy E_1 is given by the equation

$$\lambda'=\frac{1}{N\pi\sigma^2\,(1+E_2/E_1)^{\frac{1}{2}}}=\frac{l\sqrt{2}}{(1+E_2/E_1)^{\frac{1}{2}}} \qquad \text{......(7)},$$

where l is the mean free path of atoms of the gas with mean energy E_2. The mobility μ of the ions is thus given by†

l

$$\mu=\frac{0{\cdot}92el_1(1+E_2/E_1)^{\frac{1}{2}}}{(mE_1)^{\frac{1}{2}}} \qquad \text{......(8)},$$

where
$$l_1=\frac{l}{4{\cdot}4\,(\omega/E_r)^{\frac{1}{2}}}.$$

The mean fractional loss of energy F in a collision between equal elastic spheres

---

* Jeans, *Dynamical Theory of Gases*, Chap. x, p. 316.

† It is interesting to note that when the ions are in thermal equilibrium with the gas atoms equation (8) may be expressed in the form given by L. B. Loeb, *Phil. Mag.* **49**, 518 (1925), thus

$$\mu=A\left[\frac{2}{\rho\,(K-1)}\right]^{\frac{1}{2}},$$

where ρ is the density of the gas, and A is 0·25. This constant is about one half of the value given in the rigorous derivation of Langevin (*loc. cit.*) which agrees with the experimental determinations for the alkali ions in argon made by A. M. Tyndall and C. F. Powell, *Proc. roy. Soc. Lond.* A, **136**, 145 (1932). This difference is due to the adoption of the simple expression (5) for the change in the mean free path due to the polarization of the gas.

ρ, A

when the distribution of their energies about their mean energies E_1 and E_2 is Maxwellian is*

$$\tfrac{2}{3}(1-E_2/E_1).$$

The average number of collisions made per second by an ion of energy $E_1=\frac{1}{2}mu^2$ may be taken to be $\bar{u}/\lambda$, i.e. $0{\cdot}92u/\lambda$. Hence if E_1 is maintained approximately constant the average rate of loss of energy in collisions is equal to

$$0{\cdot}92FE_1u/\lambda.$$

In statistical equilibrium this loss is balanced by the energy, $e\mu Z^2$, gained per second from the electric field. Thus

$$0{\cdot}92FE_1u/\lambda=e\mu Z^2,$$

and from equations (5), (7) and (8), since $E_r=\frac{1}{2}E_2(1+k)$, then

$$k\left[\frac{1-1/k}{1+k}\right]^{\frac{1}{2}}=\frac{0{\cdot}2elZ}{E_2(\omega/E_2)^{\frac{1}{2}}} \qquad \text{......(9).}$$

Now if Z is measured in volts per centimetre and E_1 and E_2 in volts, then E_2 is $\frac{1}{27}$ V. L Also $l=L/p$, where L is the mean free path of argon atoms at a pressure of 1 mm. The value of σ deduced† from the coefficient of viscosity is $2{\cdot}96\times10^{-8}$ cm. giving $L=7\times10^{-3}$ cm. Again, since $(k-1)$ for argon is $5{\cdot}29\times10^{-4}$ it follows that B is $7{\cdot}48\times10^{-43}$, so that ω is $2{\cdot}3\times10^{-13}$ ergs and E_2 is 6×10^{-14} ergs; thus $(\omega/E_2)^{\frac{1}{2}}$ is 2. Hence equation (9) becomes

$$\left[\frac{k(k-1)}{k+1}\right]^{\frac{1}{2}}=1{\cdot}88\times10^{-2}\frac{Z}{p} \qquad \text{......(10).}$$

The values of the ratio Z/p corresponding to various values of k obtained from this equation are given in the table, where they may be compared with the experimental results.

§ 8. CONCLUSIONS

It has been indicated in § 6 that the experimental values of k may be smaller than the actual values owing to the action of impurities, but for the highest values of Z/p the influence of impurities is not so important. Furthermore, the calculated values of Z/p were obtained on the assumption of the predominating influence of the polarization of the gas atoms, so that they must be regarded as upper limits to these values.

It will be seen then that although a general agreement obtains between the observed and calculated values when Z/p is less than about 80, there is a difference between them which increases when the ratio becomes higher. This difference may be due to the neglect of two considerations: (i) the mean free path of the positive ions at the velocities considered may be shorter than that deduced by equations (5) and (7) from the kinetic theory of gases; and (ii) exchange processes occurring in some of the collisions tend to diminish the mean energy of agitation of the ions. Now the mean free path of positive ions in argon has been found, at certain velocities, to be very different from that deduced on the kinetic theory of gases; an effect which is similar to the Townsend-Ramsauer effect for slow electrons. Thus Thomson* found a variation of the mean free path of protons in argon, the minimum occurring at the same velocity as with electrons; while Ramsauer and Beeck† have shown that the cross-sections of argon atoms in collision with the alkali ions increased rapidly when the energy of the ions was reduced to the order of 1 V. Such variations of the mean free path of the ions would cause corresponding variations in the mean energies of the ions.

---

* M. Cravath, *Phys. Rev.* **36**, 248 (1930).

† H. R. Hassé and W. R. Cook, *Phil. Mag.* **3**, 977 (1927).

Again, Hassé and Cook‡ have shown that the process of electron exchange reduces the mobility of positive ions by an amount which depends on the probability of a transfer occurring in any collision with an atom. Since the energy acquired by the ions in moving through the gas is determined by their velocity in the direction of the electric force, exchange processes probably tend to diminish the mean energy of the ions.

It was not possible to determine the relative importance of these considerations in the experiments described in this paper, and the investigations are being extended to measure the coefficients of ionization by positive ions of argon under strong electric forces.

§ 9. ACKNOWLEDGMENT

In conclusion I wish to thank Prof. E. J. Evans for extending the facilities of his laboratory and for his kind interest.

---

* G. P. Thomson, *Phil. Mag.* **2**, 1076 (1926).
† C. Ramsauer and O. Beeck, *Ann. Phys.*, Lpz., **87**, 1 (1928).
‡ H. R. Hassé and W. R. Cook, *Phil. Mag.* **12**, 554 (1931).

538.522:517.788

THE COMPUTATION OF THE INTEGRALS REQUIRED IN MUTUAL-INDUCTANCE CALCULATIONS

By N. F. ASTBURY, M.A., Electricity Department,
National Physical Laboratory

Communicated by Dr L. Hartshorn, August 21, 1934. *Read in title, November* 16, 1934.

ABSTRACT. A method, based on the reduction of classical formulae, whereby the numerical computation of complete elliptic integrals of the third kind can be considerably simplified in certain cases is suggested. A particular case considered in detail is that which arises in the calculation of the mutual inductance of a helix and a coaxial circle. A considerably simplified formula, well adapted for numerical working, is deduced, together with an extremely simple result for the case when the helix and circle have the same radius.

---

§ 1. INTRODUCTION

The need for non-laborious methods of computing elliptic integrals arises in several physical problems. L. V. King⁽¹⁾ has given a very detailed account of the various devices by which elliptic integrals can be reduced to forms which admit of easy numerical computation, and, in addition to many classical formulae, he has given new series suitable for the direct numerical computation of complete elliptic integrals of the third kind. In certain cases, however, which are of interest in the calculation of coefficients of mutual inductance, it appeared to the present writer that much labour could be saved in the computation of the integrals of the third kind by the direct reduction of classical formulae to simpler forms. A new expression is derived which is well adapted for numerical working, and its application to Viriamu Jones' formula⁽²⁾ for the mutual inductance of a helix and a coaxial circle is considered.

§ 2. REDUCTION OF CLASSICAL FORMULAE

$F(\phi, k)$, $E(\phi, k)$, $\Pi(\phi, k, n)$, k

Denoting by $F(\phi, k)$, $E(\phi, k)$ and $\Pi(\phi, k, n)$ respectively the incomplete elliptic integrals of the first, second and third kinds respectively to modulus k, we have

$$\left.\begin{aligned} F(\phi, k) &= \int_0^\phi d\phi/\Delta \\ E(\phi, k) &= \int_0^\phi \Delta\, d\phi \\ \Pi(\phi, k, n) &= \int_0^\phi d\phi/\{(1 + n\sin^2\phi)\,\Delta\} \end{aligned}\right\} \qquad \ldots\ldots(1),$$

Δ

where $\Delta = \sqrt{(1 - k^2 \sin^2\phi)}$.

The angle ϕ is the ***amplitude***, and n is the ***parameter*** of the third integral. The *complete integrals*, which are usually denoted by the letters K, E and Π respectively, are obtained by putting ϕ equal to $\pi/2$.

K

Methods for the evaluation of these integrals based on transformation of the moduli and amplitudes were discovered independently by Landen and by Gauss. Each method depends on the formation of a *scale of arithmetico-geometrical means*. Such a scale is formed as follows. Start with two positive numbers a_0, b_0, such that $a_0 > b_0$, and form successively the following quantities:

a_0, b_0

$$\left.\begin{array}{lll} & c_0 = \sqrt{(a_0{}^2 - b_0{}^2)} & \\ a_1 = \tfrac{1}{2}(a_0 + b_0) & b_1 = \sqrt{(a_0 b_0)} & c_1 = \tfrac{1}{2}(a_0 - b_0) \\ a_2 = \tfrac{1}{2}(a_1 + b_1) & b_2 = \sqrt{(a_1 b_1)} & c_2 = \tfrac{1}{2}(a_1 - b_1) \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ a_n = \tfrac{1}{2}(a_{n-1} + b_{n-1}) & b_n = \sqrt{(a_{n-1} . b_{n-1})} & c_n = \tfrac{1}{2}(a_{n-1} - b_{n-1}) \end{array}\right\} \quad \ldots(2).$$

a_n, b_n, c_n

The a's and b's rapidly tend to the same limit, which we may denote by $a_n = M(a_0, b_0)$, while the c's become zero in the limit. It is easy to see that, if p is any number, then

M, p

$$p . M(a_0, b_0) = M(pa_0, pb_0) \quad \ldots\ldots(3).$$

Landen's transformation is equivalent to forming the arithmetico-geometrical mean scale $a_0 = 1$, $b_0 = \sqrt{(1 - k^2)}$, $c_0 = k$, together with a trigonometrical recurrence formula for the amplitudes (starting with $\phi_0 = \phi$)

$$\tan(\phi_{r+1} - \phi_r) = (b_r/a_r) \tan \phi_r \quad \ldots\ldots(4),$$

and just as the a's and b's tend to a limit, so the ϕ's derived from equation (4) continually approach a value which is equal to 2^r multiplied by a definite magnitude. In other words, $\phi_r/2^r$ approaches a limit which we may denote by Φ.

Φ

We then have the following results⁽³⁾ for incomplete integrals of the first and second kinds:

$$F(\phi, k) = \Phi/a_n \quad \ldots\ldots(5),$$

$$E(\phi, k) = (E/K)\, F(\phi, k) + \sum_{r=1} c_r \sin \phi_r \quad \ldots\ldots(6),$$

and for the complete integrals of the first and second kinds

$$K = \pi/2a_n \quad \ldots\ldots(7),$$

$$(K - E)/K = \tfrac{1}{2} \sum_{r=0} 2^r c_r{}^2 \quad \ldots\ldots(8).$$

We now proceed to apply these results to the reduction of the classical formulae for complete elliptic integrals of the third kind. Legendre showed that, in general, such an integral could be expressed in terms of complete and incomplete integrals of the first and second kinds, the result depending on the form of the parameter n of the third integral. In what follows we shall be concerned with the case when n is negative and particularly when n is such that $1 > |n| > k^2$, although in passing it is worth while noting, on account of its simplicity, the case in which $|n| < k^2$.

In this latter case, Legendre writes $n = -k^2 \sin^2 \theta$ and deduces the result⁽⁴⁾

$$\Pi - K = \frac{\tan \theta}{\sqrt{(1 - k^2 \sin^2 \theta)}} [K . E(\theta, k) - E . F(\theta, k)] \quad \ldots\ldots(9).$$

Forming the arithmetico-geometrical mean scale $a_0=1$, $b_0=\sqrt{(1-k^2)}$, $c_0=k$, using the recurrence formula (4), and applying results (5)–(8), we can easily throw this equation into the very simple form

$$\Pi-K=\frac{\pi\tan\theta}{2a_n\sqrt{(1-k^2\sin^2\theta)}}\sum_{r=1}c_r\sin\phi_r \qquad \text{......(10)},$$

which is very suitable for numerical work. If we work to 8-figure accuracy, the arithmetico-geometrical mean scale converges in four terms, while only two or three terms are required in the summation.

Turning now to the former case, in which we are directly interested from the point of view of mutual-inductance calculation, we note that Legendre writes

$$n=-1+k'^2\sin^2\phi_0' \qquad \text{......(11)},$$

k'

in which $k'=\sqrt{(1-k^2)}$ and is called the *complementary modulus*. He derives the result(5)

$$\frac{k'^2\sin\phi_0'\cos\phi_0'}{\Delta\,(k'\,\phi_0')}(\Pi-K)=\pi/2+(K-E)\,F\,(\phi_0'\,k')-K.E\,(\phi_0'\,k') \qquad \text{...(12)}.$$

Form now the arithmetico-geometrical mean scale

$$a_0=1,\quad b_0=k',\quad c_0=\sqrt{(a_0^2-b_0^2)}=k$$

and its *complementary scale*

$$a_0'=1,\quad b_0'=c_0=k,\quad c_0'=k',$$

and in association with the complementary scale use the recurrence formula (4) starting with ϕ_0' as given by equation (11). Denoting by dashed letters all quantities related to the complementary scale and recurrence formula, we may, using results (5) to (8), write the right-hand side of equation (12) as

Φ'

$$\pi\left[\tfrac{1}{2}+\left(\tfrac{1}{2}\sum_{r=0}2^r c_r^2+\tfrac{1}{2}\sum_{r=0}2^r c_r'^2-1\right)\Phi'/(2a_n a_n')-(1/2a_n)\sum_{r=1}c_r'\sin\phi_r'\right] \qquad \text{......(13)}.$$

K', E'

Denoting by dashed letters, K', E', complete integrals of the first and second kinds respectively to modulus k', we have(6)

$$EK'+E'K-KK'=\pi/2 \qquad \text{......(14)},$$

which may be written

$$(K-E)/K+(K'-E')/K'-1=-\pi/2KK' \qquad \text{......(15)},$$

which, by equations (7) and (8), becomes

$$\tfrac{1}{2}\sum_{r=0}2^r c_r^2+\tfrac{1}{2}\sum_{r=0}2^r c_r'^2-1=-2a_n a_n'/\pi \qquad \text{......(16)},$$

enabling us to substitute for the coefficient of Φ' in equation (13), which now becomes

$$\pi\left[\tfrac{1}{2}-\Phi'/\pi-\frac{1}{2a_n}\sum_{r=1}c_r'\sin\phi_r'\right] \qquad \text{......(17)}.$$

Finally, if we denote by Ψ' the complement, *in degrees*, of Φ', equation (12) reduces to the form

$$(\Pi - K)/\pi = \{\Delta\,(k'\,\phi_0')/(2k'^2 \sin\phi_0' \cos\phi_0')\}\left\{\Psi'/90 - \frac{1}{a_n}\sum_{r=1} c_r' \sin\phi_r'\right\}$$

$$= \{\sqrt{(-n)}/(k'^2 \sin 2\phi_0')\}\left\{\Psi'/90 - \frac{1}{a_n}\sum_{r=1} c_r' \sin\phi_r'\right\} \qquad \text{......(18).}$$

This formula is very convenient for numerical work, comparing favorably with King's direct formula

$$\Pi/\pi = \frac{\cos(2\psi_1 - \psi_0)}{a_n k'^2 \sin 2\psi_0}\left\{a_n \sin\phi_n + 2\sum_{r=0}(2^{r+1} - 1)\, c_{r+2}\frac{\tan(2\psi_{r+2} - \psi_{r+1})}{\cos(2\psi_{r+3} - \psi_{r+2})}\right\} \qquad \text{......(19),}$$

in which the arithmetico-geometrical mean scale $a_0 = 1$, $b_0 = k'$, $c_0 = k$ is used and the ψ's are derived from the recurrence formula

$$\sin(2\psi_{r+1} - \psi_r) = (b_r/a_r)\sin\psi_r \qquad \text{......(19.1),}$$

ψ_0 being given by

$$n = -k^2/(1 - k'^2 \sin^2\psi_0) \qquad \text{......(19.2).}$$

The process for numerical computation by formula (18) may now be briefly recapitulated as follows: (i) Construct the arithmetico-geometrical mean scale $a_0 = 1$, $b_0 = k'$, $c_0 = k$ and its complementary scale $a_0' = 1$, $b_0' = k$, $c_0' = k'$; (ii) derive Φ' or $\phi_n'/2^n$ from the recurrence formula $\tan(\phi'_{r+1} - \phi_r') = (b_r'/a_r')\tan\phi_r'$ starting with ϕ_0' as given by $n = -1 + k'^2 \sin^2\phi_0'$, and finally finding the complement Ψ' of Φ'; (iii) effect the summation $\sum_{r=1} c_r' \sin\phi_r'$.

In general, when we are working to 8-figure accuracy, the first arithmetico-geometrical mean scale will converge in four terms and the complementary scale in two or three. The amplitude scale and the summation will converge in two or three terms, the third term in the latter having perhaps only two or three significant figures.

The advantages of formula (18) lie in its comparatively simple form, in the considerably reduced work required with 8-figure trigonometrical tables (the use of which rapidly becomes tedious) and in the very simple summation which is required. These more than outweigh the work involved in the formation of a second arithmetico-geometrical mean scale. In the numerical example given below only nine references to trigonometrical tables were required, as against seventeen when King's formula (19) was used, the whole computation occupying less than half the time.

The following figures, taken from an actual practical case, illustrate the use of formula (18):

$$k^2 = 0{\cdot}8282\,6099, \qquad k'^2 = 0{\cdot}1717\,3907,$$

$$n = -0{\cdot}9653\,3002, \qquad 1 + n = 0{\cdot}0346\,6998.$$

Arithmetico-geometrical mean scale: $a_0 = 1$, $b_0 = k'$, $c_0 = k$.

r	a	b	c
0	1·0000 0000	0·4144 1412	0·9100 8845
1	0·7072 0706	0·6437 5005	0·2927 9294
2	0·6754 7856	0·6747 3297	0·0317 2850
3	0·6751 0576	0·6751 0565	0·0003 7280
4	0·6751 0570	0·6751 0570	0·0000 0006

Arithmetico-geometrical mean scale: $a_0'=1$, $b_0'=k$, $c_0'=k'$.
Amplitude scale: $\Phi'=\phi_n'/2^n=(205^\circ\ 6'\ 39{\cdot}5'')/2^3=25^\circ\ 38'\ 19{\cdot}94''$.

r	a'	b'	c'	ϕ'
0	1·0000 0000	0·9100 8845	0·4144 1412	26° 41′ 57·06″
1	0·9550 4422	0·9539 8556	0·0449 5578	51° 17′ 35·71″
2	0·9545 1489	0·9545 1474	0·0005 2933	102° 33′ 19·77″
3	0·9545 1482	0·9545 1482	0·0000 0008	205° 6′ 39·54″

Hence $$\Psi=64^\circ\ 21'\ 40{\cdot}06'' \quad\text{and}\quad \Psi/90=0{\cdot}7151\ 2364,$$

$$\frac{1}{a_n}\sum_{r=1} c_r'\sin\phi_r'=\frac{0{\cdot}0350\ 8155+0{\cdot}0005\ 1667-0{\cdot}0000\ 0003}{0{\cdot}6751\ 0570}$$
$$=0{\cdot}0527\ 2980;$$

$$\therefore\ \Psi/90-\frac{1}{a_n}\sum_{r=1} c_r'\sin\phi_r'=0{\cdot}6623\ 9384.$$

Also $$\sqrt{(-n)}/(k'^2\sin 2\phi_0')=7{\cdot}1262\ 523,$$

so that $$(\Pi-K)/\pi=0{\cdot}6623\ 9384\times 7{\cdot}1262\ 523=4{\cdot}7203\ 856.$$

§ 3. APPLICATION TO MUTUAL INDUCTANCE CALCULATION

M
N, x, B, A

J. Viriamu Jones[(2)] has given the following formula for the mutual inductance M between a helix of N turns, length x and radius B, and a coaxial circle of radius A in the plane of one end:

$$M=2\pi^2N(A+B)\left\{\frac{g}{k}\frac{K-E}{\pi}+\frac{kg'^2}{g}.\frac{K-\Pi}{\pi}\right\} \qquad \text{......(20)},$$

in which the modulus k of the elliptic integrals is given by

$$k^2=4AB/\{(A+B)^2+x^2\} \qquad \text{......(20.1)},$$

and the parameter n of the third integral by

$$n=-g^2=-4AB/(A+B)^2 \qquad \text{......(20.2)}.$$

In addition $$k^2+k'^2=1=g^2+g'^2 \qquad \text{......(20.3)}.$$

Let us now write $$\left.\begin{aligned}R_1^2&=(A+B)^2+x^2\\ R_2^2&=(A-B)^2+x^2\end{aligned}\right\} \qquad \text{......(20.4)},$$

so that $$k'=R_2/R_1.$$

It is easy to see in this case that equation (18) reduces to the simple form

$$(\Pi-K)/\pi=\frac{R_1}{2g'x}\left\{\frac{\Psi}{90}-\frac{1}{a_n}\sum_{r=1} c_r'\sin\phi_r'\right\} \qquad \text{......(21)}.$$

We note now that $(A+B)\,g/k=R_1$. Bearing in mind the homogeneity relation (3), we may now form the arithmetico-geometrical mean scales

$$a_0=R_1,\qquad b_0=R_2,\qquad c_0=\sqrt{(R_1^2-R_2^2)}=\sqrt{(4AB)},$$

and $$a_0'=R_1,\qquad b_0'=\sqrt{(4AB)},\qquad c_0'=R_2,$$

together with the recurrence formula (for deriving Ψ)

$$\tan(\phi'_{r+1}-\phi_r')=(b_r'/a_r')\tan\phi_r',$$

starting with $\phi_0'=\arc\sin(g'/k')$.

Denoting by a_n the limit to which the first scale converges, and by S_1 and S_2 respectively the summations $\sum\limits_{r=0} 2^r c_r^2$ and $\sum\limits_{r=1} c_r' \sin\phi_r'$, and applying results (8), (18) and (21) to (20), we get a_n, S_1, S_2

$$M=2\pi^2 N\left[\frac{S_1}{4a_n}-\frac{A^2-B^2}{2x}\left\{\frac{\Psi}{90}-\frac{S_2}{a_n}\right\}\right]$$

$$=\pi^2 N\left[\frac{1}{2}\frac{S_1}{a_n}-\frac{A^2-B^2}{x}\left\{\frac{\Psi}{90}-\frac{S_2}{a_n}\right\}\right] \qquad \text{......(22)},$$

which is a form well adapted to numerical working.

The three dimensions A, B and x required for the problem being given, the process of calculation may be summarized as follows: (i) Find R_1 or $\sqrt{\{(A+B)^2+x^2\}}$ and R_2 or $\sqrt{\{(A-B)^2+x^2\}}$; (ii) construct the arithmetico-geometrical mean scales $a_0=R_1$, $b_0=R_2$, $c_0=\sqrt{(4AB)}$ and $a_0'=R_1$, $b_0'=\sqrt{(4AB)}$, $c_0'=R_2$; (iii) from the first scale derive the limit a_n and from the second scale and recurrence formula (4) derive Φ' and Ψ starting with $\phi_0'=\arc\sin(g'/k')=\arc\sin\{R_1(A-B)/R_2(A+B)\}$; (iv) effect the summations S_1 and S_2.

Finally, it is worth noting the case when $A=B$. Equation (22) then reduces to the very simple form

$$M=\pi^2 N S_1/2a_n \qquad \text{......(23)}.$$

This formula, which is analogous to King's formula(7) for coaxial circles, gives the mutual inductance of a helix of N turns and a single turn of the same radius in the plane of one end.

It should be mentioned that Campbell(8), Smith(9) and Dye(10) all appear to have used Legendre's result (12) in the evaluation of Jones's formula, but the simplified form (22) derived above does not seem to have been known.

§ 4. ACKNOWLEDGMENTS

The writer is obliged to Dr L. Hartshorn and Mr T. Smith, F.R.S. for their criticism of the MS.

REFERENCES

(1) L. V. King. *On the Direct Numerical Calculation of Elliptic Functions and Integrals* (C.U.P. 1924).
(2) J. V. Jones. *Proc. Roy. Soc.* A, **63**, 198 (1898).
(3) King. *Loc. cit.* p. 8.
(4) Legendre. *Fonctions Elliptiques*, **1**, 148.
(5) Legendre. *Ibid.* **1**, 151.
(6) Cayley. *Elliptic Functions* (2nd edition), p. 48.
(7) L. V. King. *Proc. Roy. Soc.* A, **100**, 63 (1921).
(8) A. Campbell. *Proc. Roy. Soc.* A, **87**, 391 (1912).
(9) F. E. Smith. *Philos. Trans.* A, **214**, 211 (1912).
(10) D. W. Dye. *Proc. Roy. Soc.* A, **101**, 315 (1922).

Further accounts of the use of arithmetico-geometrical mean series in inductance calculations will be found in papers by L. V. King and by F. W. Grover, *Phil. Mag.* Series 7, **15**, 1097 *et seq.* (June 1933).

519.281

ON OBSERVATIONS OF POINTS CONNECTED BY A LINEAR RELATION

BY MISS BERYL M. DENT, M.SC., Research Department,
Metropolitan-Vickers Electrical Co. Ltd.

Communicated by H. R. Hassé, July 10, 1934. *Read in title, November* 16, 1934.

ABSTRACT. The problem of drawing the best straight line through a set of observed points is solved by a method shorter than those previously published. It is essential for the complete solution of the problem to obtain the most probable value of the ratio of the precision constants of the two observed sets of quantities and a new method is given for finding this ratio. Expressions for the errors in the position and inclination of the line are derived and a numerical example is added.

§ 1. INTRODUCTION

THE problem of drawing the best straight line through a set of observed points is usually treated in text books on the assumption (often tacit) that only one of the coordinates is liable to error. The more general case in which both coordinates are liable to error has been dealt with by Pearson*, Stewart† and Uhler‡. Pearson lays down an arbitrary criterion for a good fit, Stewart gives the solution for the case in which the precision constants for the two coordinates are equal, and Uhler obtains the solution when the two precision constants are not equal. The general solution is however dependent on a knowledge of the ratio of the precision constants, and no indication is given of any method of estimating the values of the precision constants from the given data if these consist, as they usually do, of a series of single observations of different points. Moreover, no consideration of the errors in the position and inclination of the line are given. In this paper both these matters are dealt with. The solution of the problem when the line is constrained to pass through a given point follows without further analysis.

A paper by W. R. Cook§ gives a solution of the problem of curve-fitting by means of least squares, including a consideration of the errors and a determination of the most probable value of the ratio of the precision constants. His method, however, is only applicable to curves represented by equations of the second and higher degrees, and not to a straight line.

---

* Pearson, K., *Phil. Mag.* 6, **2**, 559 (1901).
† Stewart, R. Meldrum, *Phil. Mag.* 6, **40**, 217 (1920).
‡ Uhler, H. S., *J. opt. Soc. Amer.* **7**, 1043 (1923).
§ Cook, W. R., *Phil. Mag.* 7, **12**, 1025 (1931).

§ 2. STATEMENT OF THE PROBLEM

Given a set of S observed points $(X_1, Y_1), (X_2, Y_2) \dots (X_s, Y_s)$, whose true positions are known from theoretical considerations to lie on a straight line, the problem is to find the best straight line through them when both the X's and the Y's are liable to error.

$S, (X_r, Y_r)$

It will be assumed, to start with, that repeated observations of any one point may be made, and that it is possible to keep either coordinate constant and make repeated observations of the other. The repeated observations allow error laws for both coordinates to be formulated. It will be assumed that the error laws for all the X's are the same and that those for all the Y's are the same, but different from those for the X's. In a later paragraph a method of solution will be given for cases where these assumptions are not justified.

We shall now solve the problem of finding the straight line on which the points (x_r, y_r) lie, when the x_r and y_r are chosen so as to make it most probable that the points (x_r, y_r) are the true positions of the observed points (X_r, Y_r). The points (x_r, y_r) will be referred to as the best* points.

(x_r, y_r)

§ 3. SOLUTION. GENERAL CASE

Let $\Phi(\Delta)$ represent the frequency distribution of errors of observation in the X's and $\Phi'(\Delta)$ represent the frequency distribution of errors of observation in the Y's. $\Phi(\Delta)$ and $\Phi'(\Delta)$ are assumed to be normalized, i.e.

$\Phi(\Delta)$

$\Phi'(\Delta)$

$$\int_{-\infty}^{+\infty} \Phi(\Delta)\, d\Delta = \int_{-\infty}^{+\infty} \Phi'(\Delta)\, d\Delta = 1.$$

Then the probability that the best position of the point observed as (X_r, Y_r) is (x_r, y_r) is equal to

$$\Phi(x_r - X_r)\, \Phi'(y_r - Y_r)\, \delta x_r \delta y_r$$

and the probability that the best positions of all the observed points are

$$(x_1, y_1), (x_2, y_2) \dots (x_s, y_s)$$

---

* The word *best* has been used throughout this paper as equivalent to *most probable*. Exception may reasonably be taken to the use of the word *true* in this connection. Strictly speaking, we cannot hope to find the true positions of the observed points, we can only make a guess at their true positions. In other words, we can find those positions which, when all the given data are taken into consideration, are most probable for the observed points. Obviously, if further data come into our possession, such as a detection of systematic error, the best values found when these data also are taken into consideration will be different from those previously found. We have then made a better guess at the true values.

The admission that we cannot hope to find the true positions of the observed points, but only their most probable positions, enables us to avoid the bugbear of inverse probability.

Let the frequency-distribution diagram of repeated observations on a quantity X_r have a maximum at X_r^m and let the errors be the deviations from X_r^m. Then X_r^m is the *best* or most probable value of X_r. If $\Phi(X_r^m - X_r)$ represents the frequency-distribution of the errors, the probability that X_r^m shall be observed as lying between X_r and $X_r + \delta X_r$ is $\Phi(X_r^m - X_r)\, \delta X_r$ and $\Phi(X_r^m - X_r)\, \delta X_r^m$ is the probability that the point observed as X_r shall have a *best*, or most probable, value lying between X_r^m and $X_r^m + \delta X_r^m$. This latter statement is rigorously correct, and no assumption such as "*a priori*, all values of X_r^m are equally likely" is necessary.

X_r^m

is equal to

P
$$\prod_r \Phi(x_r - X_r)\,\Phi'(y_r - Y_r)\,\delta x_r \delta y_r = P,$$

say.

P is therefore a function of the 2*s* unknown quantities x_r, y_r. By making it a maximum, subject to the condition that the points (x_r, y_r) are collinear, we shall obtain the parameters of the required line, and also, if we desire them, the values of all the coordinates x_r and y_r.

§4. SOLUTION. GAUSSIAN ERROR FUNCTIONS

h
Let
$$\Phi(\Delta) = \frac{h}{\sqrt{\pi}}\, e^{-h^2\Delta^2}$$

h′
and
$$\Phi'(\Delta) = \frac{h'}{\sqrt{\pi}}\, e^{-h'^2\Delta^2},$$

then we have to make the expression

$$\prod_r \frac{hh'}{\pi} \exp\left[-\{h^2(x_r - X_r)^2 + h'^2(y_r - Y_r)^2\}\right] \delta x_r \delta y_r$$

a maximum, subject to the conditions

m, c
$$y_r = mx_r + c, \qquad r = 1, 2 \ldots s.$$

We must therefore make

$$h^2 \sum_{r=1}^{s} (x_r - X_r)^2 + h'^2 \sum_{r=1}^{s} (y_r - Y_r)^2$$

a minimum, subject to the conditions

$$y_r = mx_r + c.$$

Using the method of Lagrangian multipliers, we form the expression

$F(x_1 \ldots \lambda_r)$
λ_0, λ_r
$$F(x_1, x_2 \ldots x_s, y_1, y_2 \ldots y_s, m, c, \lambda_r)$$
$$\equiv \lambda_0 [h^2 \Sigma (x_r - X_r)^2 + h'^2 \Sigma (y_r - Y_r)^2] + \Sigma \lambda_r (y_r - mx_r - c) = 0,$$

$$\frac{\partial F}{\partial x_r} = \lambda_0\, 2h^2 (x_r - X_r) - m\lambda_r = 0, \qquad r = 1, 2 \ldots s \qquad \ldots\ldots(1),$$

$$\frac{\partial F}{\partial y_r} = \lambda_0\, 2h'^2 (y_r - Y_r) + \lambda_r = 0, \qquad r = 1, 2 \ldots s \qquad \ldots\ldots(2),$$

$$\frac{\partial F}{\partial m} = -\Sigma \lambda_r x_r = 0 \qquad \ldots\ldots(3),$$

$$\frac{\partial F}{\partial c} = -\Sigma \lambda_r = 0 \qquad \ldots\ldots(4),$$

$$\frac{\partial F}{\partial \lambda_r} = y_r - mx_r - c = 0, \qquad r = 1, 2 \ldots s \qquad \ldots\ldots(5).$$

These $3s + 2$ equations suffice to determine the $3s + 2$ unknown quantities

$$x_r,\ y_r,\ m,\ c,\ \lambda_r.$$

By summing equation (1) over all values of r and using (4), we obtain

$$\Sigma\,(x_r - X_r) = 0,$$

and similarly

$$\Sigma\,(y_r - Y_r) = 0,$$

i.e. the centroid of the observed points coincides with the centroid of the best points. Hence the required line passes through the centroid $(\bar{X}, \bar{Y})$ of the observed points. Change the origin to the point $(\bar{X}, \bar{Y})$ and let the new coordinates be denoted by dashed letters, so that $x_r' = x_r - \bar{X}$, and so on.

$(\bar{X}, \bar{Y})$

x_r', y_r'

Then equations (1) to (5) are replaced by

$$\lambda_0\, 2h^2\,(x_r' - X_r') - m\lambda_r = 0, \qquad r = 1, 2 \dots s \qquad \text{......(6)},$$

X_r'

$$\lambda_0\, 2h'^2\,(y_r' - Y_r') + \lambda_r = 0, \qquad r = 1, 2 \dots s \qquad \text{......(7)},$$

Y_r'

$$\Sigma\lambda_r x_r' = 0 \qquad \text{......(8)},$$

$$y_r' - m x_r' = 0, \qquad r = 1, 2 \dots s \qquad \text{......(9)}.$$

Eliminating y_r' from equation (7) by means of (9), we have

$$2\lambda_0\, h'^2\,(m x_r' - Y_r') + \lambda_r = 0 \qquad \text{......(10)}.$$

From equation (6)

$$\frac{\lambda_r}{\lambda_0} = \frac{1}{m}\, 2h^2\,(x_r' - X_r') \qquad \text{......(11)}.$$

Substituting from equation (11) in (10) and (8), we have

$$h'^2\,(m x_r' - Y_r') + (h^2/m)\,(x_r' - X_r') = 0 \qquad \text{......(12)}$$

and

$$\Sigma\,(x_r' - X_r')\,x_r' = 0 \qquad \text{......(13)}.$$

From equation (12)

$$x_r' = \frac{m h'^2\, Y_r' + h^2 X_r'}{m^2 h'^2 + h^2} \qquad \text{......(14)}.$$

On substitution in equation (13)

$$\Sigma\left(\frac{m h'^2\, Y_r' + h^2 X_r'}{m^2 h'^2 + h^2}\right)^2 - \Sigma\left(\frac{m h'^2\, Y_r' + h^2 X_r'}{m^2 h'^2 + h^2}\right) X_r' = 0$$

or

$$m^2 h'^2 \Sigma X_r' Y_r' - m\,(h'^2 \Sigma Y_r'^2 - h^2 \Sigma X_r'^2) - h^2 \Sigma X_r' Y_r' = 0 \qquad \text{......(15)}.$$

Write $h/h' = k$. Then equation (15) becomes

k

$$m^2 \Sigma X_r' Y_r' - m\,(\Sigma Y_r'^2 - k^2\, \Sigma X_r'^2) - k^2 \Sigma X_r' Y_r' = 0 \qquad \text{......(16)}.$$

A convenient method of solving this quadratic is as follows:

$$\frac{2mk}{k^2 - m^2} = \frac{2k \Sigma X_r' Y_r'}{k^2 \Sigma X_r'^2 - \Sigma Y_r'^2},$$

from equation (16). Now write

$$\tan\phi = (1/k)\tan\theta = m/k \qquad \text{......(17)}.$$

ϕ, θ

Then

$$\tan 2\phi = \frac{2m/k}{1 - m^2/k^2} = \frac{2mk}{k^2 - m^2}.$$

Therefore

$$\tan 2\phi = 2k\, \Sigma\, X_r' Y_r' / (k^2\, \Sigma\, X_r'^2 - \Sigma\, Y_r'^2) \qquad \text{......(18)}.$$

ϕ may therefore be found first from equation (18) and then m from equation (17). The position and inclination of the required line are therefore completely determined. The coordinates x_r', y_r' may also be found from equations (14) and (9). Of the two solutions for m, the correct one is that which makes

$$h^2\Sigma(x_r'-X_r')^2+h'^2\Sigma(y_r'-Y_r')^2$$

a minimum.

If the error laws have been determined by repeated observations of one of the points (X_r, Y_r) the most probable position (X_r^m, Y_r^m) of this point will have been found. We therefore choose this point as origin instead of the centroid of the observed points, and proceed as from equation (6). The value of m is still given by equations (17) and (18), but X', Y' now refer to (X_r^m, Y_r^m) as origin.

§ 5. SPECIAL CASES

(*a*) *The case in which $h=h'$.* When the precision constants for X and Y are equal, i.e. when $h=h'$, the expressions are considerably simplified. Equations (17) and (18) give

$$\tan\phi=\tan\theta=m$$

and

$$\tan 2\theta=2(\Sigma X_r' Y_r')/(\Sigma X_r'^2-\Sigma Y_r'^2) \qquad \text{......(19)}.$$

The ambiguity as to $\pi/2$ is resolved by making $\Sigma\{(x_r'-X_r')^2+(y_r'-Y_r')^2\}$ a minimum.

From equations (6) and (7) we have, when $h=h'$,[*]

$$\frac{y_r'-Y_r'}{x_r'-X_r'}=-\frac{1}{m},$$

so that the best points are the feet of the perpendiculars from the observed points to the required line, and the required line is that which makes the sum of the squares of the perpendiculars to it from the observed points a minimum.

(*b*) *The case in which $h\to\infty$, $h'\to\infty$.* If the observed values of X_r may be regarded as free from error, i.e. if $h\to\infty$, equation (15) becomes

$$m=(\Sigma X_r' Y_r')/(\Sigma X_r'^2) \qquad \text{......(20)}.$$

Similarly, if $h'\to\infty$,

$$m=(\Sigma Y_r'^2)/(\Sigma X_r' Y_r') \qquad \text{......(21)}.$$

§ 6. CONSIDERATION OF ERRORS

$$\bar{X}=(\Sigma X_r)/s, \qquad \bar{Y}=(\Sigma Y_r)/s.$$

Following Eddington[†], we shall express the errors in the centroid of the observed points and the inclination of the required line in terms of "range per risk."

μ_r (*a*) *Centroid of observed points.* The probability of an error μ_r in X_r is

$$\frac{h}{\sqrt{\pi}}e^{-h^2\mu_r^2}\,d\mu_r.$$

---

[*] In general $\frac{y_r'-Y_r'}{x_r'-X_r'}=\frac{h^2}{h'^2}\left(-\frac{1}{m}\right)$, so that all the lines joining the best to the observed points are parallel. This property holds good when the true points lie on any curve or surface, and it has previously been noted by Deming, *Phil. Mag.* **11**, 146, 1931, and Uhler, *loc. cit.*

[†] *Proc. Phys. Soc. Lond.* **45**, 271 (1933).

Therefore the probability of an error $s^{-1}\Sigma\mu_r$ in $\overline{X}$ is

$$(h^s/\pi^{s/2})\exp(-h^2\Sigma\mu_r^2)\,d\mu_1\ldots d\mu_s \qquad \ldots\ldots(22),$$

and the probability that the error in $\overline{X}$ has a value between ϵ_1 and ϵ_2 is the integral of equation (22) taken over the field of integration for which $s^{-1}\Sigma\mu_r$ lies between ϵ_1 and ϵ_2. ϵ_1, ϵ_2

This probability is therefore*

$$=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{\epsilon_1}^{\epsilon_2}\int_{-\infty}^{+\infty}\ldots\left\{\exp i\theta\left(\tau-\frac{1}{s}\Sigma\mu_r\right)\right\}(h^s/\pi^{s/2})\{\exp(-h^2\Sigma\mu_r^2)\}\times d\mu_1\ldots d\mu_s\,d\tau\,d\theta,$$

and on performing the integrations we find that the probability that the error in $\overline{X}$ lies between μ and $\mu+d\mu$ is $\psi(\mu)\,d\mu$, where $\psi(\mu)$

$$\psi(\mu)=\frac{h}{\sqrt{\pi}}\sqrt{s}\,.\,e^{-\mu^2h^2s} \qquad \ldots\ldots(23).$$

Similarly the probability that the error in $\overline{Y}$ lies between μ and $\mu+d\mu$ is $\psi'(\mu)\,d\mu$, where $\psi'(\mu)$

$$\psi'(\mu)=\frac{h'}{\sqrt{\pi}}\sqrt{s}\,.\,e^{-\mu^2h'^2s}.$$

Hence the probability that the modulus of the error in $\overline{X}$ is not greater than $\mu/\sqrt{2}$ is

$$2\int_0^{\mu/\sqrt{2}}\psi(\mu)\,d\mu \quad \text{or } p\text{, say,} \qquad \ldots\ldots(24),$$ p

and the probability that the modulus of the error in $\overline{Y}$ is not greater than $\mu/\sqrt{2}$ is

$$2\int_0^{\mu/\sqrt{2}}\psi'(\mu)\,d\mu \quad \text{or } p'\text{, say,}$$ p'

so that the probability that the best position of the centroid shall not be farther than μ from $(\overline{X}, \overline{Y})$ is pp'. I therefore take a risk of 1 in $1/(1-pp')$ of being wrong by asserting that the best position of the centroid is not farther than μ from $(\overline{X}, \overline{Y})$.

(*b*) *Inclination of the required line.* The error in m can readily be deduced in any special case from the error in $\tan 2\phi$ by means of the equation $m=k\tan\phi$. If the error in X_r' be μ_r and the error in Y_r' be ν_r, then ν_r

$$\tan 2\phi=\frac{2k\,\Sigma(X_r'+\mu_r)(Y_r'+\nu_r)}{k^2\Sigma(X_r'+\mu_r)^2-\Sigma(Y_r'+\nu_r)^2}.$$

Expanding by Taylor's theorem, we find that the error in $\tan 2\phi$ is

$$\frac{2k\,\Sigma\,\mu_rY_r'}{(k^2\Sigma X_r'^2-\Sigma Y_r'^2)}-\frac{4k\,\Sigma X_r'Y_r'\,\Sigma\,\mu_rX_r'}{(k^2\Sigma X_r'^2-\Sigma Y_r'^2)^2}$$
$$+\frac{2k\,\Sigma\,\nu_rX_r'}{(k^2\Sigma X_r'^2-\Sigma Y_r'^2)}-\frac{4k\,\Sigma X_r'Y_r'\,\Sigma\,\nu_rY_r'}{(k^2\Sigma X_r'^2-\Sigma Y_r'^2)^2}.$$

Writing $A=2k/(k^2\Sigma X_r'^2-\Sigma Y_r'^2)$ and $B=4k\,\Sigma X_r'Y_r'/(k^2\Sigma X_r'^2-\Sigma Y_r'^2)^2$, we have that the error in $\tan 2\phi$ A, B

$$=A\Sigma\,\mu_rY_r'-k^2B\,\Sigma\,\mu_rX_r'+A\,\Sigma\,\nu_rX_r'+B\,\Sigma\,\nu_rY_r'.$$

---

* Whittaker and Robinson, *Calculus of Observations*, pp. 168 and 175 (1929).

Now the probability that there is an error μ_r in X_r' is

$$\frac{h}{\sqrt{\pi}}\, e^{-h^2\mu_r^2}\, d\mu_r,$$

and the probability that errors μ_r and ν_r occur simultaneously in the values X_r' and Y_r' is

$$(hh'/\pi)^s \exp\left[-\{h^2\Sigma\,\mu_r^2+h'^2\,\Sigma\,\nu_r^2\}\right].$$

The probability that the error in tan 2ϕ has a value between ϵ_1 and ϵ_2 is the integral of this expression taken over the field of integration for which

$$A\,\Sigma\,\mu_r Y_r'-k^2\,B\,\Sigma\,\mu_r X_r'+A\Sigma\,\nu_r X_r'+B\Sigma\,\nu_r Y_r'$$

lies between ϵ_1 and ϵ_2.

By the method of the preceding paragraph it can be shown that the probability
$\chi(\mu)$ that the error in tan 2ϕ lies between μ and $\mu+d\mu$ is $\chi(\mu)\,d\mu$, where

$$\chi(\mu)=\frac{1}{\sqrt{\pi}}\,\frac{e^{-\mu^2/E^2}}{E} \qquad \text{......(25)},$$

in which

E
$$E^2=h^{-2}\,\Sigma\,(AY_r'-k^2\,BX_r')^2+h'^{-2}\,\Sigma\,(AX_r'+BY_r')^2$$
$$=h^{-2}\,\{(A^2+k^2B^2)\,(\Sigma\,Y_r'^2+k^2\,\Sigma\,X_r'^2)\}.$$

The probability that the modulus of the error in tan 2ϕ is not greater than ϵ is therefore

q
$$2\int_0^{\epsilon}\chi(\mu)\,d\mu \quad \text{or} \quad q, \text{ say}, \qquad \text{......(26)}.$$

I therefore take a risk of 1 in $1/(1-q)$ of being wrong by asserting that the best value of tan 2ϕ does not differ by more than ϵ from the value

$$2k\,(\Sigma X_r'\,Y_r')/(k^2\,\Sigma X_r'^2-\Sigma Y_r'^2).$$

§ 7. MOST PROBABLE VALUES OF THE PRECISION CONSTANTS

The foregoing analysis may be applied directly to the problem provided that the precision constants for the X and Y coordinates can be determined. If this is not possible, we must try to use the given data to determine their most probable values.

In the expression

$$\prod_r\,(hh'/\pi)\exp\left[-\{h^2\,(x_r-X_r)^2+h'^2\,(y_r-Y_r)^2\}\right]\,\delta x_r\,\delta y_r,$$

which is equal to the probability that the best positions of all the observed points are $(x_1, y_1)\ldots(x_s, y_s)$, h and h' must now be regarded as variables.

Substituting $y_r=mx_r+c$ in this expression, we write

$F'(x_1\ldots c)$
$$F'(x_1\ldots x_s, h, h', m, c)\equiv(hh'/\pi)^s\exp\left[-\{h^2\,\Sigma\,(x_r-X_r)^2+h'^2\,\Sigma\,(y_r-Y_r)^2\}\right].$$

For maximum probability, F' is a maximum and therefore

$$\frac{\partial F'}{\partial x_r}=\frac{\partial F'}{\partial h}=\frac{\partial F'}{\partial h'}=\frac{\partial F'}{\partial m}=\frac{\partial F'}{\partial c}=0.$$

These equations give

$$h^2(x_r - X_r) + h'^2(mx_r + c - Y_r)\,m = 0, \qquad r = 1, 2 \ldots s \qquad \ldots\ldots(27),$$

$$s = 2h^2\,\Sigma\,(x_r - X_r)^2 \qquad \ldots\ldots(28),$$

$$s = 2h'^2\,\Sigma\,(mx_r + c - Y_r)^2 \qquad \ldots\ldots(29),$$

$$\Sigma\,(mx_r + c - Y_r)\,x_r = 0 \qquad \ldots\ldots(30),$$

$$\Sigma\,(mx_r + c - Y_r) = 0 \qquad \ldots\ldots(31).$$

From equations (31) and (27) the centroid theorem follows as before.

Changing the origin to the centroid of the observed points and denoting the new coordinates by dashed letters, we therefore have the following equations to determine h, h', x_r', m:

$$h^2(x_r' - X_r') + h'^2(mx_r' - Y_r')\,m = 0, \qquad r = 1, 2 \ldots s \qquad \ldots\ldots(32),$$

$$s = 2h^2\,\Sigma\,(x_r' - X_r')^2 \qquad \ldots\ldots(33),$$

$$s = 2h'^2\,\Sigma\,(mx_r' - Y_r')^2 \qquad \ldots\ldots(34),$$

$$\Sigma x_r'\,(mx_r' - Y_r') = 0 \qquad \ldots\ldots(35).$$

From equation (32) by squaring:

$$h^4(x_r' - X_r')^2 = h'^4(mx_r' - Y_r')^2\,m^2, \qquad r = 1, 2 \ldots s$$

and summing over all values of r, we have

$$\frac{h^4}{h'^4} = m^2\,\frac{\Sigma\,(mx_r' - Y_r')^2}{\Sigma\,(x_r' - X_r')^2}.$$

From equations (33) and (34)

$$\frac{h^2}{h'^2} = \frac{\Sigma\,(mx_r' - Y_r')^2}{\Sigma\,(x_r' - X_r')^2}.$$

Hence

$$h^2/h'^2 = m^2 \qquad \ldots\ldots(36).$$

Substituting in equations (32) and (35) and solving for m and x_r', we have

$$x_r' = (mX_r' + Y_r')/2m \qquad \ldots\ldots(37)$$

and

$$m^2 = (\Sigma\,Y_r'^2)/(\Sigma\,X_r'^2) \qquad \ldots\ldots(38).$$

The most probable values of h and h' may now be determined separately from equations (33) and (34).

§ 8. ERRORS CORRESPONDING TO THE MOST PROBABLE VALUES OF h AND h'

(*a*) *Centroid of observed points.* The analysis of § 6 (*a*) is unaltered; we simply use the most probable values of h and h' in the expressions for $\psi(\mu)$ and $\psi'(\mu)$.

(*b*) *Inclination of the required line.* When k has its most probable value,

$$m = (\Sigma\,Y_r'^2)^{\frac{1}{2}}\,(\Sigma\,X_r'^2)^{-\frac{1}{2}}.$$

If μ_r and ν_r are the errors in X_r' and Y_r' respectively,

$$\begin{aligned} m &= \{\Sigma\,(Y_r' + \nu_r)^2\}^{\frac{1}{2}}\,\{\Sigma\,(X_r' + \mu_r)^2\}^{-\frac{1}{2}} \\ &= (\Sigma\,Y_r'^2)^{\frac{1}{2}}\,(\Sigma\,X_r'^2)^{-\frac{1}{2}} - (\Sigma\,Y_r'^2)^{\frac{1}{2}}\,(\Sigma\,X_r'^2)^{-\frac{3}{2}}\,\Sigma\,X_r'\mu_r \\ &\quad + (\Sigma\,Y_r'^2)^{-\frac{1}{2}}\,(\Sigma\,X_r'^2)^{-\frac{1}{2}}\,\Sigma\,Y_r'\nu_r, \text{ etc.} \end{aligned}$$

The error in m is therefore $A' \Sigma X_r' \mu_r + B' \Sigma Y_r' \nu_r$, where

A'
$$A' = -(\Sigma Y_r'^2)^{\frac{1}{2}} (\Sigma X_r'^2)^{-\frac{3}{2}},$$

B'
$$B' = (\Sigma Y_r'^2)^{-\frac{1}{2}} (\Sigma X_r'^2)^{-\frac{1}{2}}.$$

By the same method as before it may be deduced that the probability that the

$\chi'(\mu)$ error in m lies between μ and $\mu + d\mu$ is $\chi'(\mu)\, d\mu$, where

E'
$$\chi'(\mu) = \frac{1}{\sqrt{\pi}} \frac{e^{-\mu^2/E'^2}}{E'} \qquad \text{......(39)},$$

and
$$E'^2 = \frac{1}{h^2} \Sigma (A' X_r')^2 + \frac{1}{h'^2} \Sigma (B' Y_r')^2 = m^2 \left(\frac{1}{h^2} + \frac{1}{h'^2}\right).$$

The probability that the modulus of the error in $\tan\theta$ is not greater than ϵ is therefore equal to

$$2 \int_0^\epsilon \chi'(\mu)\, d\mu \quad \text{or} \quad q', \text{ say} \qquad \text{......(40)}.$$

I therefore take a risk of 1 in $1/(1-q')$ of being wrong by asserting that the best value of m does not differ by more than ϵ from the value $(\Sigma Y_r'^2)^{\frac{1}{2}} (\Sigma X_r'^2)^{-\frac{1}{2}}$.

§ 9. NUMERICAL EXAMPLE

As an application let us take the four-point example quoted by Uhler*, Stewart† and Merriman‡.

The observed points are

$$X = 0{\cdot}4, \quad 0{\cdot}6, \quad 0{\cdot}8, \quad 0{\cdot}9;$$
$$Y = 0{\cdot}5, \quad 0{\cdot}8, \quad 1{\cdot}0, \quad 1{\cdot}2.$$

We find first $\bar{X}$ and $\bar{Y}$:

$$\bar{X} = 2{\cdot}7/4 = 0{\cdot}675;$$
$$\bar{Y} = 3{\cdot}5/4 = 0{\cdot}875.$$

The coordinates referred to the centroid as origin are:

$$X' = -0{\cdot}275, \quad -0{\cdot}075, \quad +0{\cdot}125, \quad +0{\cdot}225;$$
$$Y' = -0{\cdot}375, \quad -0{\cdot}075, \quad +0{\cdot}125, \quad +0{\cdot}325.$$

Since h and h' are not given, we must use equation (38) to find the most probable value of m:

$$m^2 = (\Sigma Y_r'^2)/(\Sigma X_r'^2) = 0{\cdot}2675/0{\cdot}1475 = 1{\cdot}813,$$
$$m = 1{\cdot}347, \quad \theta = \tan^{-1} m = 53^\circ\, 25'.$$

The line is shown in figure 1.

If we assume, as Stewart does, that the precision constants h and h' are equal, we have to use the formula (19), viz.

$$\tan 2\theta = \frac{2\Sigma X_r' Y_r'}{\Sigma X_r'^2 - \Sigma Y_r'^2} = \frac{2 \times 0{\cdot}1975}{0{\cdot}1475 - 0{\cdot}2675} = -3{\cdot}29.$$

Therefore
$$2\theta = 106^\circ\, 54', \qquad \theta = 53^\circ\, 27'.$$

---

* Uhler, *J. opt. Soc. Amer.* 7, 1057 (1923).

† Stewart, *Phil. Mag.* 6, **40**, 223 (1920).

‡ Merriman, *Textbook on the Method of Least Squares.* 1885 edn. p. 127. Art. 107.

The two values to be compared with Stewart's are

(1) $\tan\theta = 1{\cdot}349$,

(2) the intercept on the y-axis $= \overline{Y} - m\overline{X} = 0{\cdot}875 - 1{\cdot}349 \times 0{\cdot}675 = -0{\cdot}035$,

both of which are exactly equal to Stewart's second (final) approximation.

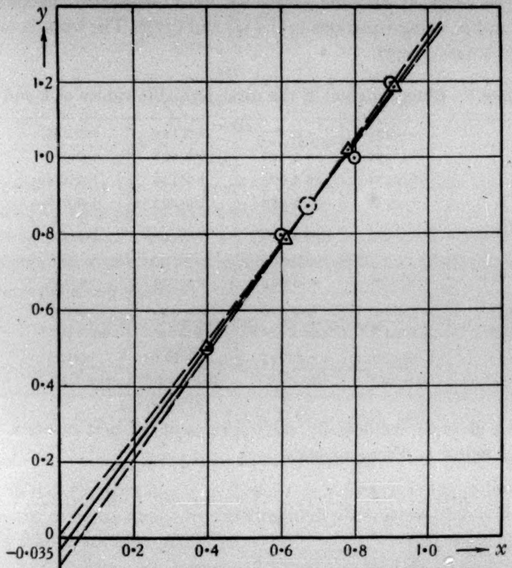


Figure 1. Straight line drawn through four observed points. The dotted circle and lines indicate the errors for a risk of 1 in 100. ⊙, observed. ▽, calculated.

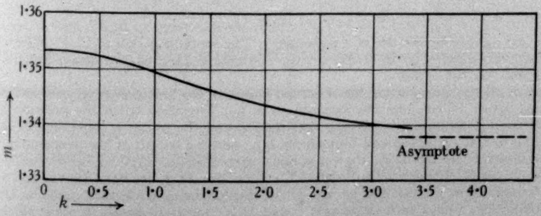


Figure 2. m as a function of k, equation (16)

We note that the two values of m obtained by making $k = 1$ or $1{\cdot}35$ differ by less than 1 per cent. It is now of interest to find the effect on m of allowing k to range over its whole series of possible values, from 0 to ∞. The extreme values are found to be $m=1{\cdot}353$ corresponding to $k=0$, and $m=1{\cdot}338$ corresponding to $k=\infty$. The curve m against k is shown in figure 2. In this particular case, then, we can obtain the value of m correct to 1 per cent, without even knowing the ratio of the precision constants.

Errors. In order to find the errors we must determine the most probable values of h and h', using equations (33), (34) and (37)*. The work is set out in the following table ($m=1{\cdot}347$).

Table 1. Determination of the most probable values of h and h'

X'	$-0{\cdot}275$	$-0{\cdot}075$	$+0{\cdot}125$	$+0{\cdot}225$	Σ
mX'	$-0{\cdot}370$	$-0{\cdot}101$	$+0{\cdot}168$	$+0{\cdot}303$	—
Y'	$-0{\cdot}375$	$-0{\cdot}075$	$+0{\cdot}125$	$+0{\cdot}325$	—
$mX'+Y'$	$-0{\cdot}745$	$-0{\cdot}176$	$+0{\cdot}293$	$+0{\cdot}628$	—
$\dfrac{mX'+Y'}{2}=y'$	$-0{\cdot}372$	$-0{\cdot}088$	$+0{\cdot}146_5$	$+0{\cdot}314$	—
$\dfrac{mX'+Y'}{2m}=x'$	$-0{\cdot}277$	$-0{\cdot}065$	$+0{\cdot}109$	$+0{\cdot}233$	—
$(x_r'-X_r')$	$-0{\cdot}002$	$+0{\cdot}010$	$-0{\cdot}016$	$+0{\cdot}008$	—
$(x_r'-X_r')^2$	4×10^{-6}	100×10^{-6}	256×10^{-6}	64×10^{-6}	42×10^{-5}
$(y_r'-Y_r')$	$+0{\cdot}003$	$-0{\cdot}013$	$+0{\cdot}021_5$	$-0{\cdot}011$	—
$(y_r'-Y_r')^2$	9×10^{-6}	169×10^{-6}	462×10^{-6}	121×10^{-6}	76×10^{-5}

This table gives us, incidentally, the positions of the best points x_r', y_r'. These are shown in figure 1. From equation (33)

$$h^2=\frac{4}{2\Sigma\,(x_r'-X_r')^2}=\frac{2}{42\times10^{-5}}=4760,\qquad h=69{\cdot}0;$$

and from equation (34)

$$h'^2=\frac{4}{2\Sigma\,(y_r'-Y_r')^2}=\frac{2}{76\times10^{-5}}=2630,\qquad h'=51{\cdot}3.$$

As a check on the work we note that $h^2/h'^2=4{\cdot}760/2{\cdot}630=1{\cdot}811=(1{\cdot}346)^2$.

---

* The sums $\Sigma\,(x_r-X_r)^2$ and $\Sigma\,(y_r-Y_r)^2$, which are necessary to find h and h' separately, may be obtained very simply by the use of a planimeter. The method is due to C. F. Merriman (see the discussion following a paper by E. O. Waters on "Graphical Methods for Least-Square Problems," *A.S.M.E. Trans.* **51**, 201 (1929)).

The slope of all the lines joining the observed points to the best points when h and h' have their most probable values is $-m$; see the footnote on p. 93. Therefore when the position and slope of the line have been determined, the best points may at once be found graphically. If the length of the line joining the rth observed and best points is l_r, and the length of the perpendicular from the rth observed point to the line is d_r, then it is easily seen that $(x_r-X_r)^2=l_r^2\cos^2\theta=d_r^2/4\sin^2\theta$, and $(y_r-Y_r)^2=d_r^2/4\cos^2\theta$. Circles with centres (X_r, Y_r) and radii d_r are now drawn, and a planimeter is started at some point on the line to the left of the left-hand circle and made to trace the line and to pass clockwise round each circle as it comes to it, returning finally straight along the line to the starting point. The planimeter reading will give the total area of the circles, i.e.

$$\pi\Sigma d_r^2=\pi4\sin^2\theta\Sigma\,(x_r-X_r)^2=\pi4\cos^2\theta\Sigma\,(y_r-Y_r)^2.$$

When the number of observed points is large, the amount of computation saved by this method is considerable.

(*a*) *Error in centroid.* (i) *Risk for a distance* 0·01. The probability p, that the modulus of the error in $\overline{X}$ is not greater than $0{\cdot}01/\sqrt{2}$, is *p*

$$\frac{2}{\sqrt{\pi}}\int_0^{0\cdot01/\sqrt{2}} e^{-4h^2\mu^2}\, d\,(2h\mu)$$

from equations (23) and (24).

Writing $t=2h\mu$, we have *t*

$$p=\frac{2}{\sqrt{\pi}}\int_0^{0\cdot01\times\sqrt{2}\times h} e^{-t^2}\,dt$$

$$=\frac{2}{\sqrt{\pi}}\int_0^{0\cdot97} e^{-t^2}\,dt=0{\cdot}83.$$

Similarly $$p'=\frac{2}{\sqrt{\pi}}\int_0^{0\cdot725} e^{-t^2}\,dt=0{\cdot}69.$$

The probability that the true position of the centroid shall not be farther than 0·01 from $(\overline{X}, \overline{Y})$ is $pp'=0{\cdot}83\times0{\cdot}69=0{\cdot}57$.

I therefore take a risk of 1 in $1/(1-0{\cdot}57)$, i.e. 1 in 2·3, of being wrong by asserting that the true position of the centroid is not farther than 0·01 from $(\overline{X}, \overline{Y})$.

(ii) *Distance for a risk* 1 *in* 100.

$$1/(1-pp')=100, \quad pp'=99/100.$$

Now $$p=\frac{2}{\sqrt{\pi}}\int_0^{2h\mu/\sqrt{2}} e^{-t^2}\,dt=I\,(2h\mu/\sqrt{2}), \quad \text{say}$$ *I*

and $$p'=\frac{2}{\sqrt{\pi}}\int_0^{2h'\mu/\sqrt{2}} e^{-t^2}\,dt=I\,(2h'\mu/\sqrt{2}), \quad \text{say.}$$

Therefore $$pp'=I\,(97\mu)\times I\,(72{\cdot}5\,\mu)=0{\cdot}99.$$

Hence μ is about 0·025 (shown in figure 1 as a dotted circle). I therefore take a risk of 1 in 100 of being wrong by asserting that the true position of the centroid is not farther than 0·025 from $(\overline{X}, \overline{Y})$.

(*b*) *Error in inclination.* From equations (39) and (40) the probability that the modulus of the error in m is not greater than ϵ is

$$=q'=2\int_0^{\epsilon}\chi'(\mu)\,d\mu, \quad \text{where} \quad \chi(\mu)=\frac{1}{\sqrt{\pi}}\,\frac{e^{-\mu^2/E'^2}}{E'},$$

and

$$E'^2=m^2\left(\frac{1}{h^2}+\frac{1}{h'^2}\right)=1{\cdot}813\,\frac{42+76}{2}\times10^{-5}=1{\cdot}07\times10^{-3},$$

$$E'=3{\cdot}27\times10^{-2}.$$

(1) *Risk for an error* 0·01. The probability that the modulus of the error in m is not greater than 0·01 is therefore

$$=\frac{2}{\sqrt{\pi}}\int_0^{\mu=0\cdot01} e^{-(0\cdot01)^2/E'^2}\,\frac{d\mu}{E'}$$

$$=\frac{2}{\sqrt{\pi}}\int_0^{0\cdot01/E'} e^{-t^2}\,dt=\frac{2}{\sqrt{\pi}}\int_0^{0\cdot316} e^{-t^2}\,dt$$

$$=0{\cdot}345.$$

I therefore take a risk of 1 in 1/(1 − 0·345), i.e. 1 in 1·5, of being wrong by asserting that the value of *m* differs from 1·347 by less than 0·01. In this particular case an error of ± 0·01 in *m* corresponds to an error of about ± 13′ in θ.

(2) *Error for a risk* 1 *in* 100. If we make $q' = 0{\cdot}99$, we have the equation

$$\frac{2}{\sqrt{\pi}}\int_0^{\mu/F'} e^{-t^2}\,dt = 0{\cdot}99 \text{ to determine } \mu.$$

Hence $\mu/F' = 1{\cdot}82$, therefore $\mu = 1{\cdot}82 \times 3{\cdot}27 \times 10^{-2} = 0{\cdot}06$.

Therefore I take a risk of 1 in 100 of being wrong by asserting that the error in *m* is less than 0·06. In this particular case an error of ± 0·06 in *m* corresponds to an error of about ± 1° 15′ in θ (shown in figure 1 by dotted lines).

§ 10. THE ASSUMPTION OF GAUSSIAN ERROR LAWS

The question now arises as to whether repeated observations of physical quantities actually do obey error laws of the Gaussian type. This is a point which can only be decided by experiment in each particular case. Three different cases present themselves.

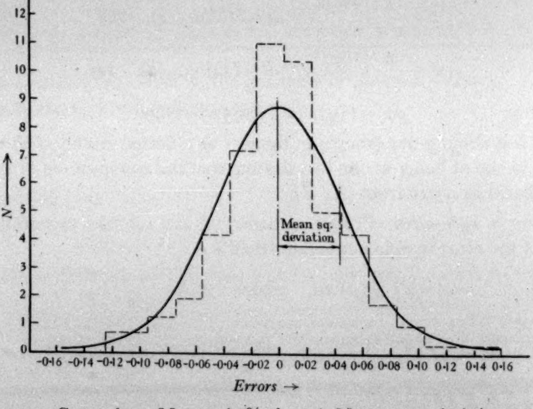


Figure 3. Copper loss. Mean, 2·62 %; *h*, 15·6; Mean-square deviation, $0{\cdot}04_5$.

(i) If repeated observations of the two coordinates of any one point can be made separately, then the results will show whether Gaussian error laws are obeyed and, if so, what are the values of the precision constants. These values will suffice for the solution of the problem dealt with in this paper, provided it can at the same time be assumed that the error laws for all the *X*'s are the same and that the error laws for all the *Y*'s are the same.

(ii) If repeated observations of any one point can be made, but it is not possible to keep one coordinate constant while making repeated observations on the other, it will still be theoretically possible to make a three-dimensional model of the frequency distribution. By taking sections along the X and Y axes in turn, the types of error law may be ascertained and their parameters determined.

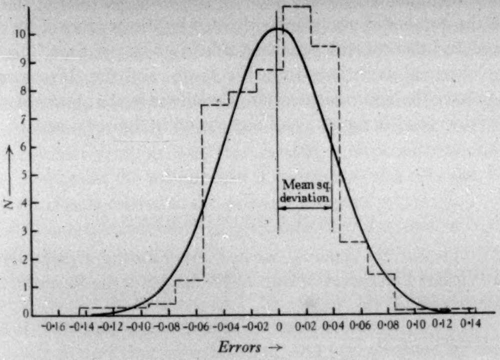


Figure 4. Impedance voltage. Mean, 3·14 %; *h*, 18·1; Mean-square deviation, 0·039.

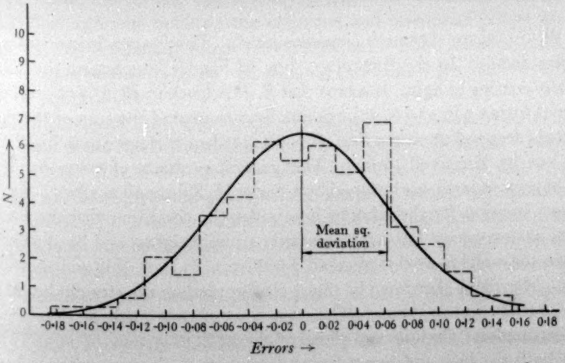


Figure 5. Iron loss. Mean, 2·11 %; *h*, 11·5; Mean-square deviation, 0·062.

(iii) If it is impossible to make any repeated observations at all, the only possible test is to examine the errors $(x_r' - X_r')$ and $(y_r' - Y_r')$ as soon as x_r' and y_r' have been determined. If a sufficiently large number of points have been observed, these errors should be distributed according to the Gaussian law, using the most probable values of the precision constants. If it is found that they are not so distributed, the method of the foregoing pages does not apply.

It may be noted here that this test should be made as a check even when repeated observations are possible.

Three examples of frequency-distribution curves of errors which do actually obey the Gaussian law are given in figures 3–5. They were obtained from the measurement of the percentage copper loss, impedance volts and iron loss of 310 transformers of identical design and made in the same workshop. In each case the ordinate N is the number of occurrences divided by the product of the total number of observations and the interval, the interval being 0·02 per cent. The broken lines represent the observed deviations from the mean, and the theoretical curves are drawn so as to have the same mean-square deviations as the observations. It is seen that the fit is fairly good in figure 3 and quite good in figures 4 and 5.

§ 11. ACKNOWLEDGMENT

I have great pleasure in thanking Mr A. P. M. Fleming, C.B.E., Director of the Metropolitan-Vickers Electrical Co. Ltd. and Manager of the Research Department, for permission to publish this work.

DISCUSSION

Dr W. Edwards Deming (*communicated*). This paper brings up a host of interesting details. In the first place, bits of history concerning the subject are continually coming to light. It seems that R. J. Adcock in 1878* first worked on the problem of fitting a line to a set of points by minimizing the sum of the squares of the normals dropped upon the line. That he made a curious slip in his algebra was pointed out by Kummell (*infra*). The general problem of curve-fitting by the method of least squares was outlined by Charles H. Kummell in 1879†. He obtained the correct solution for the straight line under the condition that the x and y coordinates have constant but not necessarily equal weights, and he showed how to determine the weights of the adjusted parameters in both this and the non-linear problem. Mansfield Merriman in 1890‡ also worked on the straight line, but made no new contribution to theory. The next advance was made by Karl Pearson in 1901 in his memoir "On lines and planes of closest fit to systems of points in space§," wherein both the best and worst-fitting lines and planes were obtained, together with ingenious and now well known formulae for the mean-square residual, the premise being that all coordinates have equal weight. Apparently unaware of earlier work, Corrado Gini in 1921|| found the best and worst-fitting lines for points whose x and y coordinates have constant but not necessarily equal weights. Gini's

---

* *Analyst* (Des Moines), **5**, 53–54 (1878).

† *Analyst*, **6**, 97–105 (1879).

‡ *Report of the United States Coast and Geodetic Survey*, p. 687 (1890).

§ *Phil. Mag.* **2**, 559–572 (1901).

|| *Metron*, **1**, No. 3, 63–82 (1921).

paper has, in turn, been overlooked by most subsequent writers, among whom I have been numbered.

Most of the investigators who have attempted to take account of errors in both x and y observed coordinates have worked on the straight line, and have recognized the impossibility of getting an exact solution without making the assumption that the x coordinates all have equal weight and the y coordinates all have equal weight. Unfortunately this assumption divests the problem of much that promised to be of importance, for it is a fact that lines obtained with various values of the ratio h/h' (in the author's notation) differ only in the squares of the residuals. Now since in practice the residuals are usually small, it turns out that the lines thus obtained usually do not differ significantly. Miss Dent's drawings afford a good illustration of this point. The non-linear relation does not exhibit this peculiarity, nor does the linear relation, save under the assumption of constant x- and constant y-weighting.

These facts are closely related to the author's observation that Mr Cook's method of estimating h/h' is applicable only to equations of the second or higher degree. Mr Cook's method breaks down in the case of the straight line only in the sense that it requires to be carried to a second approximation, which is just what one would expect in view of the fact that the ratio h/h' becomes of secondary importance when the precisions of the x and y coordinates are each constant.

It is interesting to note that equations (33) and (34) obtained by the author for estimates of h and h' differ from the usual ones in having s where one usually sees $(s-2)$. The author's formulae were based on the assumption that the horizontal and vertical line segments between the observed and calculated points are actual errors, whereas they are usually considered as residuals. The "number of degrees of freedom," as it was named by R. A. Fisher, is s if these line segments are errors, and $(s-2)$ if they are residuals.

Miss Dent recognizes the importance of testing the calculated inclination and intercept of the fitted line. From the work of Helmert, "Student," Fisher, and others, it is known that if the errors of observation are normally distributed the values of the calculated precision constants will not be so; hence the equations of §6 for testing the precision of the slope and intercept of the fitted line must be regarded as approximations which, however, improve as the number of observed points increases. Henry Schultz* has worked out a complete system for estimating the standard errors of a curve and its parameters. In this connection, Chapter 5 of R. A. Fisher's *Statistical Methods for Research Workers* should not be overlooked. The risk of 1 in $1/(1-pp')$ for an error μ in the centroid, as given by the author in §6, is more restrictive than necessary, but a full treatment would necessitate opening for discussion the whole subject of statistical tests.

Finally, a word should be said regarding the normal error law†. Even when it is not obeyed—and it probably never was—statistical deductions based on sampling from a normal parent population are not apt to lead one into wrong conclusions, provided reasonable care is exercised. But if one is interested in testing the normal law, or any other law, the approach might well be made with some objective criterion such as Karl Pearson's chi test.

---

* *J. Amer. stat. Ass.* **25**, 139–185 (1930).

† Commonly called Gaussian; see, however, footnote 8 in a paper by W. Edwards Deming and Raymond T. Birge in *Reviews of Modern Physics*, **6**, 119–161 (1934).

AUTHOR's reply. I should like to thank Dr Deming for his interesting historical notes.

In comparison with any method which requires to be carried to a second approximation, or which involves the preliminary determination of approximate values for the parameters, the method of the present paper is much simpler and shorter.

In §6 the values of h and h' are regarded as accurately known. It seems doubtful whether further approximations, involving the errors in the calculated precision constants, would be justified.

537·533·706 548·74

OXIDE FILMS ON LIQUID METALS STUDIED BY MEANS OF ELECTRON-DIFFRACTION

By R. O. JENKINS, A.R.C.S., B.Sc.,
Imperial College of Science and Technology

Received July 16, 1934. *Read November* 2, 1934.

ABSTRACT. Electron-diffraction patterns obtained by reflection from molten lead, zinc, bismuth and tin have shown that the surfaces of these metals are covered with thin oxide films. From these patterns it has been deduced that the films consist of small electron-optically flat oxide crystals resting on their [001] faces. The inner potentials and crystal structures of these oxides have been found, and they have thus been identified as definite chemical compounds. It has also been found possible to remove these films from the molten metal by means of nickel gauze and to obtain electron-diffraction patterns by transmission. The transmission patterns indicate that although the removed films are of the same chemical composition as those on the molten metal, their removal has disturbed the orientation and given the crystals an approximately Maxwellian distribution of orientations about their original direction.

§ 1. INTRODUCTORY

The work to be described in this paper was the outcome of an attempt to study the surface of liquid metals by means of electron-diffraction. The metals used, which had to be of fairly low melting-point, all acquired quite thick oxide layers when melted in air, and became oxidized to some extent even in a fairly good vacuum. During an attempt to eliminate these oxide films, curious orientations of the crystals comprising the films were observed by means of electron-diffraction. The experiments to be described consisted mainly in photographing electron-diffraction patterns by reflection from the oxide-covered surface of the molten metal, and identifying the oxide and its mode of orientation.

In addition to this, experiments were carried out by means of a new method of removing the oxide film from the surface of the metal so that that electron beam might be diffracted by transmission through the thin film. These oxide films were formed on the metal in air, while those used for reflection were formed *in vacuo*. The structure and orientation however seem to be similar in the case of the metals tried, except that the orientation was not complete in the transmission specimens.

§ 2. EXPERIMENTAL APPARATUS

The diffraction camera used for this work was one of the Thomson-Fraser type(1) with modifications designed for Prof. Thomson by C. G. Fraser of Aberdeen, and set in a horizontal plane to obtain the electron beam at grazing incidence on the liquid surface. The discharge tube had the usual aluminium disc cathode, but its other end was coned and ground to fit a corresponding ground cone on the apparatus. The joint was made vacuum tight with Apiezon M grease and cooled by a water jacket round the outside, figure 1. The electron beam was defined by the anode which consisted of a hollow aluminium cylinder with an aluminium plate round the outside at the discharge-tube end. Each end of the hollow cylinder was closed by a molybdenum disc pierced with a hole 0·15 mm. in diameter, which allowed a fine electron beam to pass from the discharge tube to the camera. Directly in front of the anode was an aluminium shutter operated from the outside of the specimen chamber through a coned, ground and greased spindle, figure 2, which also operated an interconnected razor-blade scraper just above the surface of the specimen, figure 1.

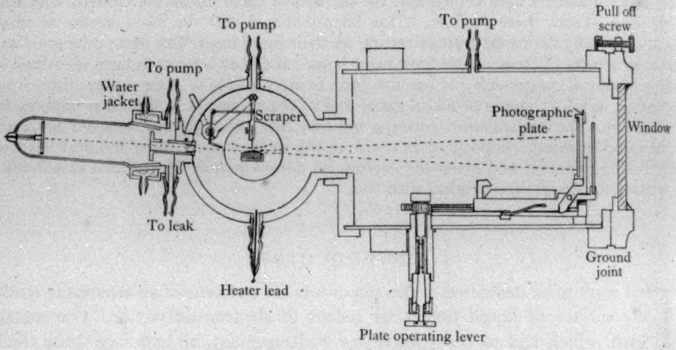


Figure 1. Diagram of apparatus.

The method of carrying the specimen may be seen from figure 2. By turning the large wheel on the outside of the specimen-chamber the specimen could be rotated in azimuth by a gear comprising two large bevel wheels, while by operating the smaller lever, a rack and pinion raised or lowered the specimen across the electron beam. Each of these movements worked independently through the two small coned and ground spindles. In addition, the angle that the specimen surface made with the beam could be altered by turning the large coned and ground joint which carried the whole specimen assembly.

The specimen was removed from the apparatus by breaking the flat ground and greased joint round the outside edge of the cylindrical specimen-chamber. This joint being nearly 2 cm. wide and about 15 cm. in diameter, considerable force was needed to break it. This was applied by means of a specially designed screw clamp after the safety ring had been removed. When the specimen was in position inside the apparatus it could be observed through a plate-glass window sealed with picein vacuum wax to the far side of the specimen-chamber.

Access to the photographic plate and its carrier was by means of the end cover of the camera, which also had a large flat ground and greased joint with drawing-off screws, figure 1. To this cover was waxed a plate-glass window through which the specimen and fluorescent screen could be observed. The plate-holder consisted of a carriage running on two rails fixed to the body of the camera and it could be moved backwards and forwards by means of a rack and pinion worked through a conical greased and ground spindle. The photographic plate was carried by the lid of a light-tight cassette, usually held down by a spring catch in a horizontal position on to the rest of the cassette so as to allow the diffracted beams to hit the fluorescent screen. When the carriage was racked towards the end of the camera, a rod in contact with the catch came up against a stop at the end of the rails and released the catch, allowing the cassette lid carrying the plate to spring up into the vertical position as shown in figure 1. After the pattern had been photographed the plate carriage could be racked back along the rails, when the two projecting lugs on the cassette lid each came in contact with a strong spring blade and closed the lid down again to the horizontal position.

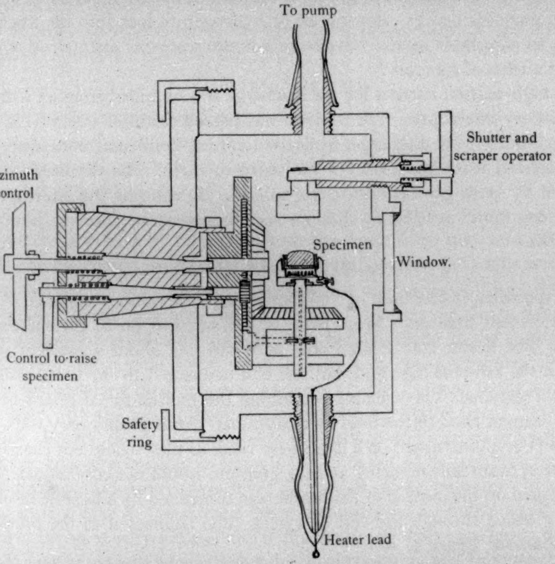


Figure 2. Section through specimen-chamber excluding razor-blade scraper.

§ 3. VACUUM AND HIGH TENSION EQUIPMENT

The whole apparatus was evacuated through wide-bore glass tubing connected through a liquid-air trap to a mercury diffusion pump backed by a Cenco-Hyvac pump. The connections to the apparatus were made by means of standard taper coned ground and greased joints, one for each of the three sections of the apparatus, figure 1. Under working conditions the pressure in the discharge tube should be about 10^{-3} mm. of mercury and that in the camera rather lower. The difference in pressure was obtained by allowing air or hydrogen to leak into the discharge-tube through an adjustable needle valve from a bottle reservoir maintained at a pressure of about 4 mm. of mercury.

The high-tension current for the discharge was obtained from an induction coil with mercury interrupter. The current was passed through a rectifying valve and water resistance to the discharge, while a smoothing-condenser of capacity $0{\cdot}004\,\mu$F. was connected from the anode of the valve to earth. The discharge voltage was measured by spark-gap between 5 cm. spheres. By running the interrupter at high speed it was found possible to obtain a very homogeneous electron beam with this equipment.

§ 4. SPECIMEN-HEATER

The specimen-heater used in conjunction with the camera consisted of a circular asbestos cement base on a brass plate which screwed on to the specimen-pillar, figure 2. The heater coil of nichrome wire carrying about 2 amperes at 30 volts rested on the asbestos base and had one end connected to the brass plate and the other to a terminal. The crucible containing the molten metal rested in a ring of asbestos cement fixed to the base by two screws. After preliminary tests, crucibles of Acheson graphite turned on a lathe were used, as they would not charge up when the electron beam hit the metal, and the graphite would not contaminate the metal. The terminal on the outside of the heater was connected by a flexible insulated lead to a wire sealed through the end of a glass tube connected to the bottom of the crystal chamber, figure 2.

§ 5. THEORY

The electron beam has a wave-length associated with it (De Broglie) and can therefore be diffracted by crystals in a similar manner to X-rays according to θ, d Bragg's law $n\lambda = 2d \sin\theta$, where θ is the angle of diffraction and d the lattice spacing of the crystal planes. Combining these two equations and the camera-dimensions, we find that

$$d = \frac{L\sqrt{150}}{r\sqrt{v}} \times 10^{-8} \text{ cm.} \qquad \ldots\ldots(1),$$

L, v, r where L is the camera-length, v the voltage of the discharge and r the distance from the diffracted spot to the undeviated spot on the photographic plate.

If the surface is electron-optically flat, so that electrons enter and leave the specimen in planes parallel to the surface, refraction takes place as a result of the difference in potential between the inside and outside of the specimen. This is

μ, ϕ

equivalent to a refractive index μ equal to $\sqrt{(1+\phi/v)}$(2), where ϕ is the inner potential. This alters all component distances of diffracted spots in a direction perpendicular to the shadow-edge. It may be shown that if the angles made with the specimen-surface by the incident and diffracted beams are equal

$$y^2 = y_0^2 - 4L^2\phi/v \qquad \text{......(2)},$$

y

y_0

where y is the observed component distance of r along a line perpendicular to the specimen-surface, and y_0 is what the distance would be if no refraction had taken place. If a rotation picture is taken and a number of orders of the same plane appear on the line perpendicular to the specimen-surface through the undeviated spot, equation (2) may be put into the form

$$y^2 = n^2 r_0^2 - 4L^2\phi/v \qquad \text{......(3)},$$

n, r_0

where n is the order number and r_0 the distance from the first-order spot to the undeviated spot, as calculated from the Bragg angle. If y^2 is plotted against n^2, the slope of the graph will give a value for r_0 which may be substituted in equation (1) to give the lattice spacing d, while the intercept will give a value of the inner potential ϕ.

The patterns obtained by reflection from the oxide films may be explained as follows. If a single crystal is placed in the beam with a cleavage face as surface, the resulting pattern on the plate consists of spots at the intersection of three diffraction curves as follows. The atoms on a line in the surface perpendicular to the electron beam give a series of equidistant straight lines perpendicular to the surface of the specimen. Even if the crystal is very small these lines will be quite sharp. The atoms lying in the surface along the beam give a series of circles. Owing to the small angle this line makes with the beam, these circles may become quite broad if the crystal is very small. Finally the planes parallel to the surface give a series of lines parallel to the surface, spaced according to Bragg's law as modified by the refraction effect. There will be a spot wherever these three diffraction curves intersect. If, however, the crystal surface may be considered flat and continuous for only a small area the second curve becomes very diffuse, and if the scattering atoms are heavy the penetration of the electrons will be small and the third condition will be relaxed to some extent. These effects have been discussed by Kirchner and Raether in some detail(3). In the patterns under consideration the crystals are small enough to make the second diffraction condition so relaxed that its effect may be ignored, and the pattern thus consists of spots at the intersection of the lines perpendicular and parallel to the shadow-edge on the photographic plate. Owing to the relaxation of the third condition these Bragg spots are elongated along the line perpendicular to the shadow-edge, so that the lines are continuous but are of greater intensity where a spot should occur. The meniscus of the surface gives the equivalent of a rotation picture and a large number of spots may thus appear, while the side lines correspondingly may be quite long. If in addition to alteration of the angle of incidence the crystal is rotated in azimuth about an axis perpendicular to its surface, the atomic distances along the line perpendicular to the electron beam will change, and a series of side lines corresponding to all the possible atomic distances in the surface will be obtained.

If the [001] plane is in the surface these distances are the lattice spacings of the [*xy*0] planes, where *x* and *y* may have any integral values. The series of spots on one particular side line due to spacing [*hj*0] will correspond to the zone of planes [*hjz*], where *z* will have any integral value.

With this theory as a basis it has been possible to evaluate two of the axes of the oxide crystals from the distances apart of the diffraction side lines. The third axis and inner potential have been obtained by applying equation (3) to the measurements on the line through the undeviated spot. From the length of the axes the oxides have been in most cases identified by comparison with X-ray crystal structures. It has thus been shown that the oxide surfaces consist of quite small crystals usually resting on their [001] faces on the liquid metal surface.

§ 6. EXPERIMENTAL RESULTS

Lead. Pure lead was melted in air in a graphite crucible and the surface was skimmed. After cooling it was placed in the apparatus, which was evacuated. The current in the heating-coil was then switched on, so as to melt the metal and maintain it at a temperature not very far above its melting-point. Owing to degassing of the heater it was necessary to wait for a few minutes till the vacuum had recovered sufficiently to allow a diffraction pattern to be taken. The specimen was then raised completely across the electron beam till the swinging razor-blade scraper could traverse the surface when the operating lever was moved. In this way a clean patch of metal was obtained. The specimen was lowered again till the electron beam could graze the surface, when the diffraction pattern could be observed on the fluorescent screen and subsequently photographed.

The complexity of the pattern suggested that the oxide had not one of the simplest forms of symmetry, figure 3. A complete analysis of the pattern was made and it was deduced that the oxide was the rhombic form of PbO known as litharge, which has axes *a*, *b*, *c* equal to 5·50 A., 4·68 A., and 5·88 A. respectively, all the small crystals being arranged with their [001] planes parallel to the surface of the specimen. The definition of the spots on the lines was bad owing to the smallness of the crystal-size and lack of penetration of the electron beam, lead being a heavy atom. The measurement of the spots on the central line was thus not very accurate, but the value of the *c* axis of the crystals calculated from these measurements agreed within 1½ per cent with the X-ray value.

As the *a* and *b* axes of the crystals were unequal, the side lines corresponding to the [*jhz*] planes would be different from those of the [*hjz*] planes, and this would make the number of lines very large. Some of the lines did not however appear owing to the form of the space group in the unit cell. From the X-ray study of litharge by the Debye-Scherrer ring method(4), the absence of the rings due to the [10*z*] [01*z*] [21*z*] [12*z*] [12*z*] [30*z*] [03*z*] [32*z*] and [23*z*] planes meant that the corresponding side lines would be missing, so that the pattern would be simplified. Table 1 gives a summary of measurements made on plates obtained with lead oxide.

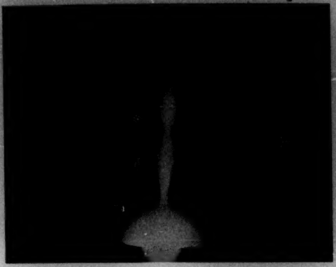


Figure 3. Reflection pattern from litharge on molten lead.

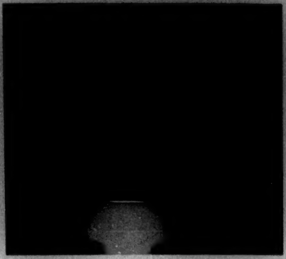


Figure 4. Reflection pattern from litharge on lead crystal.

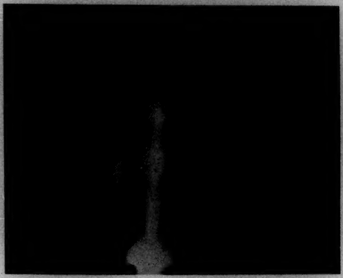


Figure 5. Reflection pattern from zinc oxide on molten zinc.

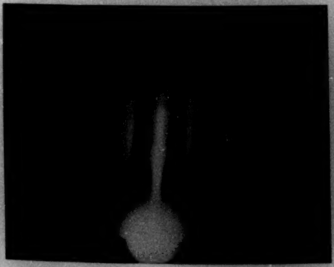


Figure 6. Reflection pattern from bismuth oxide on molten bismuth.

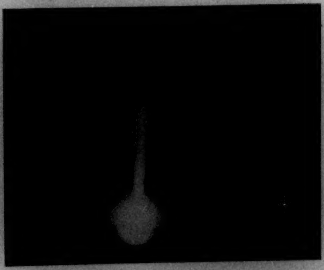


Figure 7. Reflection pattern from tin monoxide on molten tin.

In addition to the side lines tabulated, a very faint side line inside the [11z] line seemed to correspond to the forbidden [10z] series of planes.

Table 1

Indices of zone of planes causing side line	Calculated lattice spacing, $z=0$ (A.)	Observed lattice spacing, $z=0$ (A.)	Intensity
11z	3·57	3·58	Very strong
20z	2·75	2·78	Medium
02z	2·34	2·36	Medium
22z	1·79	1·79	Strong
31z	1·71	1·71	Medium
13z	1·51	1·52	Medium
40z	1·37	1·37	Very weak
33z	1·19	1·20	Medium

The mean value of the c axis was 5·97 A. and the inner potential 12·3 V.

The value for the c axis is about $1\frac{1}{2}$ per cent too high, but some of this error may be due to the thermal expansion of the crystals, which would slightly increase all the lattice spacings. Even if this effect is neglected, the value is within the limit of experimental error. The figure obtained for the inner potential is probably only accurate to about one volt.

Assuming that the small crystals comprising the layer had rotational freedom round the c axis when the metal was liquid, it was of some interest to see whether this was still the case when the metal underneath was of crystalline form, or whether any orientation took place on crystallization of the metal. This question was tested by allowing the lead to cool down very slowly by decreasing the current in the heater coil, when the lead would crystallize at the surface into a few large flat crystals from whose surfaces diffraction patterns could be obtained. It was found that on solidification the small oxide crystals on the lead had all tended to set in one direction on the top of the metal crystal, for true single-crystal diffraction pictures due to the oxide were obtained. These showed plainly only one set of equally spaced side lines with sharp spots on them, placed according to the other two diffraction conditions, figure 4. It can be seen that the second diffraction condition was obeyed rigidly, all the spots appearing on a circle. This shows that the surface must have been flat and continuous for a much larger area than in the case of the small crystals on the liquid metal. In addition to these spots there were Kikuchi lines which would also appear only with a true single crystal. It thus appears that on crystallization of the underlying metal the small oxide crystals turn so that they all are in the same azimuth, and thus make themselves into a layer of true single-crystal form on the top of the metal crystal.

Attempts were also made to reduce the oxide layer by means of atomic hydrogen, so as to leave a clean metal surface. A discharge tube of the Woods type was waxed in position through a hole in the inspection window on the far side of the specimen-chamber, and was arranged to run at a pressure of about $\frac{1}{10}$ mm. of mercury, being supplied from a bottle reservoir of hydrogen at a pressure of a few centimetres through an adjustable leak. The atomic hydrogen made by the discharge escaped through a pinhole close to the specimen-surface where it was possible that it might reduce the oxide. The oxide was not reduced but underwent rather interesting and striking changes. After the atomic-hydrogen tube had been run for about a minute and the diffraction pattern from the surface had been observed directly the discharge was cut off, two entirely new patterns were obtained. These both were due to crystals with a body-centred tetragonal lattice resting on their [001] faces, while these crystals were not electron-optically flat, no refraction effect being observed. The approximate values of the axes for these crystals were as follows: $a=4{\cdot}10$ A., $c=6{\cdot}00$ A. for one pattern, and $a=3{\cdot}85$ A., $c=13{\cdot}30$ A. for the other. Neither of these structures correspond to a known lead compound, while the second one suggests something fairly complex owing to its very large c axis.

Zinc. Pure zinc was melted in a graphite crucible and after cooling was placed in the apparatus. The procedure for melting and skimming the surface inside the apparatus was the same as that used for lead. It was found that molten zinc had a considerable tendency to vaporize, necessitating cleaning of the window and metal parts in the specimen-chamber after the taking of a diffraction photograph, and for this reason only a few plates were taken. These however gave quite definite information as to the chemical nature and orientation of the oxide surface. The film consisted of zinc-oxide (ZnO) crystals of ordinary close-packed hexagonal structure ($a=3{\cdot}22$ A., $c=5{\cdot}18$ A.) resting on their [001] faces, with their c axes perpendicular to the surface.

In the case of hexagonal crystals oriented in this manner the side lines due to the $[hjz]$ planes would correspond to lattice spacings $a\sqrt{3}/2\,(h^2+hj+j^2)^{\frac{1}{2}}$, while owing to the fact that the crystals were of close-packed structure many of the Bragg spots would not appear on the side lines. The spots in the case of zinc were rather better defined than those from lead, because the lighter zinc atoms allowed greater electron penetration and resolving power, figure 5.

The measurements for zinc oxide are given in table 2 and show quite good agreement with X-ray values, besides giving a figure for the inner potential of the oxide correct to approximately one volt.

Table 2

Indices of zone of planes causing side line	Calculated lattice spacing, $z=0$ (A.)	Observed lattice spacing, $z=0$ (A.)
$10z$	2·79	2·79
$11z$	1·61	1·62
$20z$	1·39	1·40
$21z$	1·05	1·06
$30z$	0·93	0·93
$22z$	0·80	0·81

From measurements on the central line $c=5{\cdot}27$ A. [X-ray value 5·18]. Inner potential $\phi=12{\cdot}3$ V.

The lattice spacings from the side lines and from the centre line are again slightly high, but they are just within the limits of experimental error even when thermal expansion is neglected. In addition to the side lines tabulated, a very faint one was observed inside the [10*z*] line which gave a lattice spacing for $z=0$ of approximately 3·86 A. This does not appear to bear any simple relation to the other side lines and may be caused by an oriented impurity. It may be connected with an extra ring obtained by Finch and Quarrell(5) who used electron-diffraction by transmission through an incompletely oriented film of zinc oxide. This ring gave an observed lattice spacing of 3·79 A. and was tabulated as a half order of the [102] plane of zinc oxide. It might however be the ring due to the first plane of the zone causing the side line by reflection, while the rings due to the rest of the zone of planes would be of greater radius and less intensity and might not be strong enough to appear on the plate.

Bismuth. The oxide surface on bismuth was prepared in exactly the same way as for the previous metals, the metal being used at a temperature not far above its melting-point. The resultant pattern, figure 6, showed that the oxide had a hexagonal structure with the bismuth atoms placed in the unit cell in approximately the same positions as the zinc atoms in the unit cell of zinc oxide, and that the crystals of the oxide layer were all resting on their [001] faces. The axes have been calculated from the measurements but the oxide has not been identified chemically because no crystal structures of bismuth oxides are given in Landolt and Börnstein's tables or Ewald and Herrmann's *Structurbericht*(6). As bismuth is trivalent the oxide would probably be of the formula Bi_2O_3. The other oxides of this trivalent type having a hexagonal structure are those of the rare-earth group such as cerium. The structure of compounds of this type [D 52] places the metal atoms in a close-packed hexagonal lattice with an axial ratio of 1·56 and an *a* axis of approximately 3·90 A. These figures agree very closely with those obtained for the bismuth oxide which are 1·57 for the axial ratio and 4·02 A. for the *a* axis. The suggested structure is made even more probable by the fact that the ionic diameters of the rare earths and bismuth are nearly the same.

In addition to the side lines given in table 3, two very faint lines were observed

Table 3

Indices of zone of planes causing side line	Mean observed lattice spacing $z=0$ (A.)	Corresponding value of *a* axis (A.)
10*z*	3·47	4·01
11*z*	2·02	4·04
20*z*	1·73	4·00
21*z*	1·32	4·03
30*z*	1·16	4·02
22*z*	1·01	4·04
31*z*	0·965	4·02

Mean value of $a=4·02$ A. From central line $c=6·30$ A. Hence $c/a=1·57$.
Mean value of inner potential $\phi=11·0$ V.

inside the [10z] line which did not appear to bear any simple relation with the main pattern. These again might be due to an oriented impurity as suggested in the case of zinc.

Tin. An oxide surface on pure tin was prepared as with the previous metals, the tin being used at a temperature just above its melting-point. The resultant pattern, figure 7, when analysed showed that the film consisted of the tetragonal form of SnO [$a=3{\cdot}80$ A., $c=4{\cdot}81$ A.], all the crystals resting on their [001] faces. u The tin atoms are placed in the unit cell in positions (000) and ($\frac{1}{2}\,\frac{1}{2}\,2u$), where u is an undetermined parameter. As all orders of the [001] plane appeared on the central line u cannot be $\frac{1}{4}$, which would place the tin atoms in a body-centred lattice. The side lines obtained from the tetragonal lattice corresponded to planes [hjz], where the lattice spacing for $z=0$ was $a/(h^2+j^2)^{\frac{1}{2}}$. The mean measurements for these side lines are given in table 4 and show quite good agreement with the calculated values. The value for the c axis is about 2 per cent too high, which is just within experimental error due to the diffuseness of the spots caused by lack of penetration of the electrons. The inner potential obtained from the measurements would also be correct only to about one volt.

Table 4

Indices of zone of planes causing side line	Strength of side line	Observed lattice spacing, $z=0$ (A.)	Calculated lattice spacing, $z=0$ (A.)
10z	Strong	3·80	3·80
11z	Strong	2·68	2·69
20z	Medium weak	1·90	1·90
21z	Medium	1·70	1·70
22z	Medium weak	1·35	1·34
30z	Weak	1·28	1·27
31z	Strong	1·21	1·20
32z	Strong	1·05	1·05

From measurements on central line $c=4{\cdot}91$ A.
Mean inner potential $\phi=14{\cdot}2$ V.

In addition to that already explained, an entirely different pattern also was obtained from the tin surface at a slightly higher temperature. This pattern, while apparently due to an oriented layer of character rather similar to that of the layer of SnO obtained previously, was considerably more complex and while some of its main lattice spacings have been evaluated it has not been connected with any definite oxide of tin with a particular orientation.

§ 7. OBSERVATIONS OF DIFFRACTION BY TRANSMISSION

These results having been obtained by means of the reflection arrangement, it was thought desirable to observe diffraction patterns by transmission through thin films of oxide. This has already been done for some metals with films prepared in different ways. Ponte⁽⁷⁾ obtained films of zinc oxide showing no orientation, by burning zinc in air. Bragg and Darbyshire⁽⁸⁾ obtained films by drawing a loop of wire out of the molten metal, thus leaving a film of metal across the loop. In this film were holes across which thin membranes of oxide suitable for transmission diffraction were stretched. Using this method they obtained films of SnO_2, PbO_2 and ZnO. Films of zinc oxide prepared in this way have been studied in more detail by Finch and Quarrell(5), who showed conclusively that although the diffraction patterns resembled the face-centred cubic type they were really due to the ordinary close-packed hexagonal structure of ZnO, with the intensities of the rings greatly altered by the partial selective orientation of the crystals in the film. As these facts applied to zinc oxide it seemed possible that a similar pattern might be obtained from a film of bismuth oxide, the two reflection patterns being very similar, and experiments were tried to see whether this was so.

Bismuth-oxide films. While the loop method of preparing a specimen may be easily carried out for zinc, which will wet other metals, bismuth presents a more difficult problem. Alternative methods were therefore tried, one of which has proved very satisfactory both for bismuth and for other metals. A piece of fine nickel gauze having about 25 meshes per centimetre was dipped under the freshly skimmed surface of the metal and then lifted out with a pair of tweezers, the gauze being kept nearly horizontal. This procedure allowed the metal to drain away through the gauze, leaving patches of oxide film caught in the meshes. The specimen so prepared was then fitted in a clamp in the apparatus in place of the furnace used for the previous part of the work.

The resultant diffraction pattern was in fact almost precisely similar in appearance to that obtained by Finch and Quarrell for zinc oxide. The lattice spacings for the rings were worked out and the indices allotted with the help of Hull and Davey's nomograms(9) which were also used in the same way for the tin oxide diffraction patterns. The rings fitted a hexagonal close-packed lattice of axial ratio 1·56 and an *a* axis of 3·93 A. in a condition of partial orientation so that the greatest number of crystals had their [001] planes in the plane of the specimen and none at right angles to this direction. The mean lattice spacings are given in table 5, together with

Table 5

Indices of ring	Intensity of ring	Observed lattice spacing (A.)	Calculated lattice spacing (A.)
110	Very strong	3·39	3·40
101	Strong	2·92	2·97
102	Weak	2·26	2·27
110	Very strong	2·01	1·97
103	Very weak	1·77	1·75
200	Strong	1·71	1·70
201	Medium	1·64	1·64
120	Medium	1·30	1·29
121	Medium	1·26	1·26
300	Medium	1·16	1·13
302	Weak	1·08	1·06
220	Very weak	1·00	0·99
304	Medium	0·94	0·91

the indices and intensities and calculated lattice spacings assuming the values for the lattice which gave the best fit on the nomogram.

The axial ratio from the transmission observations agrees quite closely with that obtained by reflection (transmission 1·56, reflection 1·57) while the *a* axis is about 2 per cent lower by transmission than by reflection, and some of this difference may be accounted for by thermal expansion. The transmission value of 3·93 A. is almost exactly equal to that of other members of the same structure group and is therefore probably the more correct value at ordinary temperature.

Tin-oxide films. Thin films of tin oxide were also made by the nickel-gauze method. The tin was maintained at a temperature slightly above its melting-point and the surface was skimmed before the gauze was dipped. The resultant diffraction pattern showed that the film made in this way consisted almost entirely of the monoxide SnO, the crystals being partially oriented with their [001] faces in the plane of the film as in the case of bismuth and zinc. Here again, no crystals had their *c* axes lying in the plane of the film, for no orders of the [001] planes were observed. The greatest number of crystals had their *c* axes nearly perpendicular to the plane of the film, since rings with the third index zero were greatly strengthened, figure 8. Rather similar patterns have been obtained by Steinheil(10) from specimens made by playing a flame on tin foil. In this case the orientation was not so marked, as the [002] ring appeared with moderate intensity. The calculated and observed lattice spacings of the rings together with their intensities are given in table 6.

Table 6

Indices of ring	Intensity of ring	Observed lattice spacing (A.)	Calculated lattice spacing (A.)
100	Very weak	3·75	3·80
SnO_2 110	Very weak	3·39	3·34
101	Medium	3·02	2·98
110	Very strong	2·71	2·69
111	Weak	2·36	2·35
200	Strong	1·92	1·90
201	Medium strong	1·79	1·77
120	Weak	1·70	1·70
121	Strong	1·62	1·60
202	Weak	1·51	1·49
122	Weak	1·42	1·39
220	Medium strong	1·35	1·34
221	Weak	1·31	1·29
203 301	Medium	1·24	1·22
130	Medium	1·20	1·20
123	Weak	1·18	1·17
302	Very weak	1·13	1·12
132	Medium	1·08	1·08
231 223	Medium	1·04	1·03

In addition to films consisting of the monoxide, some composed of the dioxide SnO_2 were made in a similar way with the tin at a rather higher temperature. This

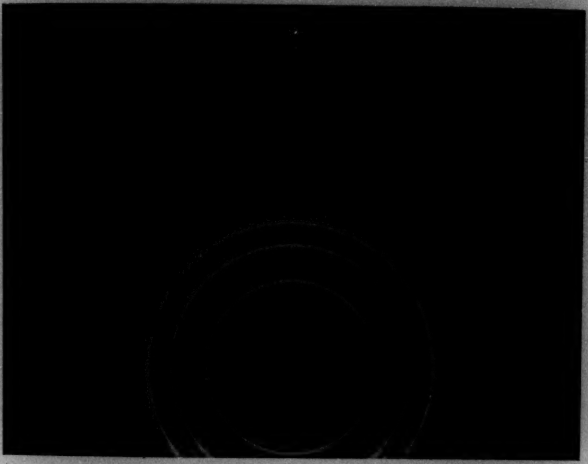


Figure 8. Transmission pattern from tin monoxide film.

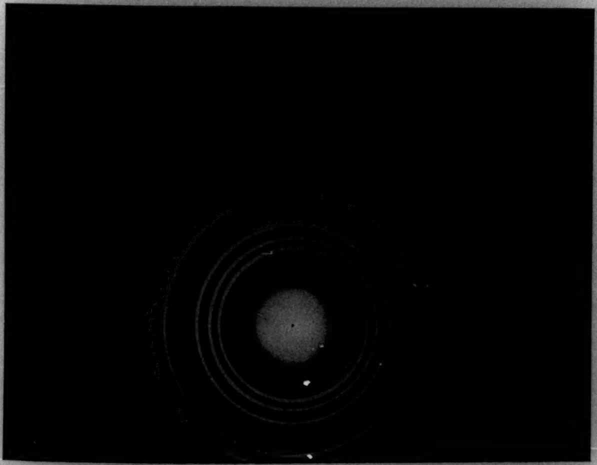


Figure 9. Transmission pattern from litharge film.

oxide layer probably started as the monoxide and changed almost entirely to the dioxide owing to the higher temperature. Similar films were also made by Steinheil by heating tin foil. The relative intensities of the specimens made by both methods agreed well and showed little sign of any orientation, since all possible rings were present. The orientation was presumably destroyed on the transition from monoxide to dioxide. The rings are given in table 7 together with their intensities and lattice spacings.

Table 7

Indices of ring	Intensity of ring	Observed lattice spacing (A.)	Calculated lattice spacing (A.)
SnO (001)	Very weak	4·77	4·81
SnO (100)	Very weak	3·75	3·80
110	Very strong	3·39	3·34
Sn (200)	Very weak	2·90	2·92
101	Medium strong	2·65	2·63
200	Strong	2·39	2·36
210	Weak	2·15	2·12
211	Very strong	1·78	1·76
220	Weak	1·69	1·67
002	Weak	1·58	1·58
310	Medium strong	1·52	1·49
301	Very weak	1·35	1·33
202	Very weak	1·31	1·31
321	Medium strong	1·23	1·21
400	Medium	1·19	1·18
222	Weak	1·16	1·15
312	Medium strong	1·10	1·09
420	Weak	1·06	1·06
103	Medium weak	1·02	1·03
402 123	Medium	0·96	0·96
510	Weak	0·93	0·93
332	Medium strong	0·92	0·92

Lead oxide. Films of lead oxide also were made with the same technique, and their diffraction patterns were photographed and analysed. These showed that the films consisted entirely of the rhombic form of PbO (litharge), the preferential orientation tending to set the crystals with their [001] planes in the plane of the film. This again caused rings due to the [00*z*] planes to be absent and those with the *z* index zero to appear with increased intensity. The rings observed by electron diffraction, figure 9, are given in table 8 together with their lattice spacings and intensities, and corresponding intensities obtained by Halla and Pawlek(4) with X-rays. A comparison of these intensities shows clearly the effect of the preferential orientation. In addition to the rings tabulated, two or three very faint rings were observed inside the [110] ring on some plates. These seemed to correspond to indices [010] [100] and $\frac{1}{2}$ [110], but may have some different origin such as an impurity with a larger unit cell.

Table 8

Indices of ring	Observed lattice spacing (A.)	Calculated lattice spacing (A.)	Observed electron-diffraction intensity	X-ray intensity
110	3·59	3·59	Medium weak	Weak
111	3·08	3·07	Strong	Strong
002	—	2·94	—	Very strong
200	2·73	2·75	Very strong	Medium
201	2·51	2·50	Very weak	Very weak
020	2·36	2·34	Very strong	Medium
112	2·21	2·28	Very weak	Very weak
202	2·03	2·02	Weak	Strong
003	—	1·96	—	Medium
022	1·87	1·85	Very weak	Strong
220	1·79	1·79	Very strong	Medium
113	1·72	1·72	Weak	Strong
311	1·65	1·64	Strong	Strong
203	—	1·60	—	Very weak
222	1·53	1·53	Weak	Strong
023	—	1·51	—	Weak
131	1·48	1·47	Strong	Strong
400	1·42	1·38	Medium weak	Very weak
114	1·37	1·36	Medium weak	Medium
223	—	1·31	—	Weak
204 313	1·29	1·29	Weak	Strong
024 402	1·25	1·25	Weak	Medium
133	—	1·21	—	Medium
330	1·18	1·19	Strong	Weak

§ 8. ACKNOWLEDGMENT

In conclusion, I should like to thank Prof. G. P. Thomson for his interest and many helpful suggestions in the course of the work.

REFERENCES

(1) Thomson and Fraser. *Proc. roy. Soc.* A, **128**, 641 (1930).
(2) Bethe. *Naturwissenschaften*, **15**, 787 (1927); **16**, 333 (1927).
(3) Kirchner and Raether. *Phys. Z.* **13**, 510 (1931).
Raether. *Z. f. Phys.* **78**, 527 (1932).
(4) Halla and Pawlek. *Z. f. Phys. Chem.* **128**, 49 (1927).
(5) Finch and Quarrell. *Proc. Phys. Soc.* **46**, 148 (1933).
(6) Ewald and Herrmann. *Structurbericht, Z. f. Krist.* Spec. (1931).
(7) Ponte. *Ann. de Physique*, **13**, 395 (1930).
(8) Bragg and Darbyshire. *Trans. Farad. Soc.* **28**, 522 (1932).
(9) Hull and Davey. *Phys. Rev.* **17**, 549 (1921).
(10) Steinheil. *Ann. der Phys.* **19**, **5**, 472 (1934).

DISCUSSION

Lord Rayleigh mentioned the fact that when sodium-nitrate crystals are deposited on a calc-spar crystal, they try to orient themselves as nearly as possible in accordance with the orientation of the latter, although the difference in structure between the two substances prevents this result from being effected completely. In the present instance, when lead-oxide crystals were formed on the surface of a large single crystal of metal did the metal atoms of the oxide remain in the positions which they had occupied in the metallic state?

Prof. G. I. Finch. In view of the reputed purity of the zinc which Dr Quarrell and I had used, we had suggested that the faint extra inner ring in our oriented ZnO patterns might be a ½ [102] of normal ZnO, but emphasized the need for caution by pointing out the fact that the supposed half-order and [102] rings arced in different directions. Since then Dr Wilman and I have found that this extra inner ring may be due to a trace of an impurity which can be almost wholly removed by skimming off the first few layers of oxide scum formed on the surface of a melt of a fresh batch of forensic zinc. The first skimming consists in the main of this impurity and is quite different in texture from a zinc-oxide scum. The transmission pattern obtained with a film of the impurity normal to the beam is shown in figure A. It has a complicated structure of a relatively low order of symmetry. One of the prominent inner rings corresponds to a spacing of 3·79 A., and this is the supposed half-order ring previously observed. Otherwise the pattern has nothing in common with either Zn, ZnO or pseudomorphic ZnO.

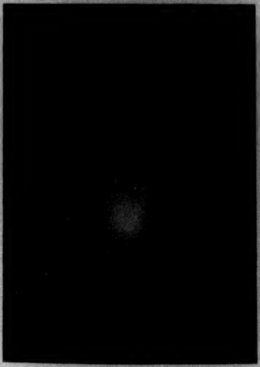


Figure A.

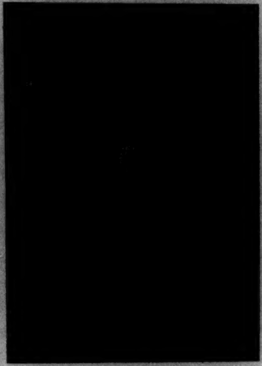


Figure B.

At first sight figure A, like many of our ZnO patterns and like those now discussed by the author, appears to be due to a random crystal array; inclination of the specimen plane to the beam by only 20° from the normal sufficed, however, to yield an arced pattern, figure B, thus proving that the crystals tended to point in a common direction. The remarkable elliptical appearance of the arced pattern is, of course, due to an illusion.

Prof. G. P. THOMSON. With reference to the single crystals of oxide which form on the surface of the lead when it solidifies, it seems to me that this is a different kind of phenomenon from those referred to by the President or Prof. Finch. In the author's experiments the small crystals of oxide are present throughout the experiment, even when the lead is melted, and the formation of a single crystal must be due to a rotation and orientation of these small pieces of crystal, brought about by the atomic forces in the metal at the moment when it solidifies. As far as I know this is a new effect, and it is of interest from the point of view of the study of crystallization.

I should also like to refer to the apparatus used by the author. This, as is stated in the paper, was designed by Mr C. G. Fraser of Aberdeen. In respect of the method of moving the specimen by means of a large coned ground joint carrying two smaller cones, figure 2, the apparatus has been imitated in others made at the Imperial College. Four such in all are now in use and some have been in operation for nearly three years. They have given very little trouble and I can recommend the mechanism as an efficient one for moving objects in a vacuum. There has been no difficulty arising from sticking of the large joints, of diameter up to 9 cm., in spite of the great pressure of the atmosphere on them.

AUTHOR's reply. In reply to Lord Rayleigh's question: the lead-oxide film on the surface of the large lead crystal was originally formed while the metal was liquid, and as the atoms in a liquid have no fixed positions there would be no relation between their positions in the lead and in the oxide. It is very probable, though the point was not investigated, that the single-crystal oxide film, caused by the rearrangement of the small oxide crystals when the underlying metal solidified, had some definite orientation with respect to the metal crystal.

Prof. G. P. Thomson has brought out the difference between the effects observed in the experiments under discussion, and phenomena observed when one substance crystallizes on another giving oriented single crystals or pseudomorphic crystals. The formation of the single-crystal lead-oxide films was caused by rotation of pre-existing small oxide crystals under the regular atomic field of the underlying metal crystal.

The ground joints to which Prof. Thomson refers have been in use for about 1½ years without any leakage or regrinding, and with only a very occasional regreasing, except in the case of the two large flat joints which needed regreasing rather more often owing to their removal for access to the photographic plate and specimen.

Prof. Finch's new diffraction pattern from the first skim of molten zinc is very interesting as it explains the diffraction ring corresponding to a lattice spacing of 3·79 A. and bears out the observation that this and the extra line obtained in the reflection patterns from molten zinc might both be due to the same oriented impurity.

535·243

THE USE OF MICROPHOTOMETRIC METHODS IN DIVIDED-BEAM SPECTROPHOTOMETRY

By D. H. FOLLETT, M.A., A.Inst.P., Adam Hilger, Ltd.

Communicated by Dr F. Simeon, August 17, 1934. *Read November* 16, 1934.

ABSTRACT. A microphotometer is described in which two photocells, connected differentially, are used to indicate the difference in transmission of the individual spectra in the pairs produced in divided-beam methods of spectrophotometry. Match points are thus found with much greater ease than by the visual method; much of the uncertainty which arises in visual matching disappears, though some remains owing to the irregularities of the photographed plate.

---

§ 1. INTRODUCTION

The use of a microphotometer for finding the match points in divided-beam methods of spectrophotometry has been suggested but so far seems never to have been put into practice. The reason for this may be that apparatus designed specially for the purpose is necessary, since the use of the ordinary form of microphotometer would be very laborious and its adaptation to this special purpose is not practicable.

§ 2. CONSTRUCTION OF APPARATUS

Figure 1 shows diagrammatically the lay-out of an instrument designed for this purpose. L contains the light-source, a 6-volt 18-watt motor head-lamp running from batteries or the mains, whichever is the more convenient. The prism P deflects the beam at right angles on to a 25-mm. microscope objective M_1; this forms an image of the filament of the lamp on the plate, which is placed with its film side uppermost on the stage S. The linear dimensions of the image are one tenth of those of the filament. The second objective M_2 forms an image of the plate on the slit B which is in the centre of the screen A. The magnification of this system is × 10. Since the filament is imaged on the plate, there is also a unit-magnification image of the filament on this slit. Each objective is furnished with a focusing motion. Two photoelectric cells of the rectifier type are mounted behind the screen, as described later. Between L and P is a lens, not shown in the diagram, which can be pushed into the beam as required; we shall refer to this as the *viewing-lens*.

The purpose of the viewing-lens is to simplify the setting of the plate on the stage. When this lens is in position, with a plate on the stage, a small patch of light, instead of an image of the filament, falls on the plate; and on the screen A appears a projected image of the spectrum under examination. The largest possible area that can be imaged on the plate is a circular patch of the same dimensions as the aperture of the microscope objectives; the position and focal length of the viewing-lens was therefore chosen so that only this area was illuminated, the image on *A* being thus made as bright as possible. When a plate taken with some such instrument as the Spekker photometer is on the stage, an image of a pair of spectra separated only by a fine dividing-line appears on the screen. The position of the plate is adjusted until the image of the dividing-line falls exactly on a horizontal line engraved on the slit and passing through its centre; under this condition the light forming the image of one spectrum falls on one photocell, and that forming the other falls on another photocell. Figure 2 shows diagrammatically the optical arrangements whereby this is achieved.

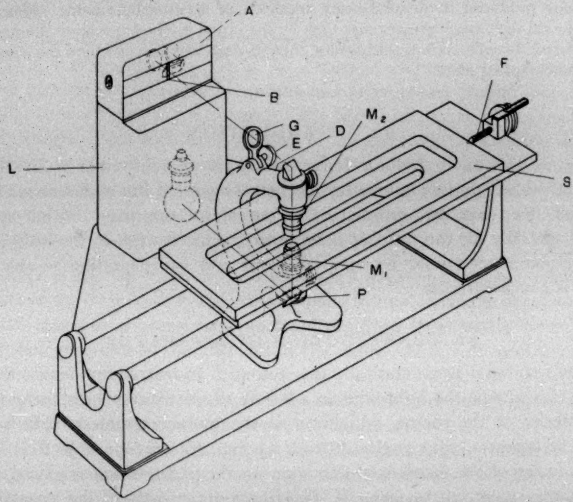


Figure 1.

The lower drawing shows an elevation, the upper a plan view of the photocell housing. The image of the dividing-line in the absorption spectrum falls on the straight line *H* which is engraved through the centre of the slit *B*, perpendicular to it. A rhomb *R* mounted behind the slit collects the light forming the image of one of the spectra and directs it on to the photocell P_1, while light forming the image of the other passes straight through on to the photocell P_2. The position of the rhomb is adjusted so that the projection of the line *H* falls along the centre line of a black band, 1 mm. wide, painted along the upper edge of the rhomb and indicated in the figure by the shaded area. This black band allows a little latitude in the positioning of the plate on the stage and prevents any difficulty over the finite width of the dividing-line.

The light falling on the cells must be divided so that the individual cells produce equal currents when an unexposed part of the plate is projected on the screen. The parallel glass plate *G*, figure 1, is provided for this adjustment; it is mounted so that its inclination to the beam can be varied between positions 22½° either side of the vertical; its thickness (about 3·5 mm.) is such that the image of the filament is displaced 1 mm. when the plate is moved from one extreme position to the other. Thus by moving the glass plate the proportion of the image of the filament falling on either side of the dividing-line can be adjusted until the amount of light falling on each cell is such as to produce equal currents in these cells.

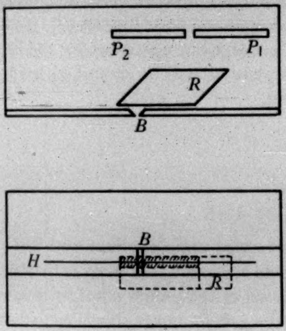


Figure 2.

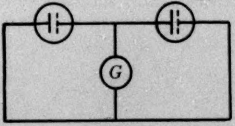


Figure 3.

The slit *B* is made with one fixed and one movable jaw; the movable one is held in position by a screw through a slotted hole, and for adjustment this screw is loosened and the jaw is slipped along in its slide. This is a sufficiently sensitive adjustment: the use of a screw is an unnecessary luxury, as the linear magnification of the image of the plate on the slit is × 10, and therefore the slit-width can be about 10 times the slit-width of the spectrograph with which the absorption photograph was taken.

The photoelectric circuit is shown in figure 3. The photoelectric cells used are made by Electrocell G.m.b.H., Berlin. The light-sensitive element can be provided as an unmounted disc 25 mm. in diameter, and this form is convenient and suitable. These cells have a high sensitivity, about 400μA. per lumen. The matching of the intensity, current characteristics for the two cells was tested by putting an Ilford wedge on the stage of the instrument and using it exactly as if it were an absorption spectrum. The beams were balanced with the light passing through the thin end of the wedge; the wedge was traversed towards the thick end, and it was observed that the balance was nowhere disturbed. The galvanometer was the A.M. type of the Cambridge Instrument Co.; its period was 6 sec., its resistance was 450 Ω., and its sensitivity about 1000 mm./μA. with the lamp and scale two metres from the mirror.

It is a great advantage to be able to use the rectifier type of photocell, as its current is very easily measured with the galvanometer; the emission type would involve the use of an amplifier which is far less convenient than the galvanometer. In order to obtain the utmost sensitivity, however, it is necessary to form an image of the source on the slit, and since the spectrum cannot then be seen on the screen, the addition of the viewing-lens already described becomes necessary.

The sensitivity of the instrument is such that the deflection obtained through a clear plate on one beam only is about 240 mm. with a slit 0·3 mm. wide while the lamp is run at 7 volts. This slit is not unduly wide, for in absorption work the spectrograph slit may permissibly be wider than in emission work. The deflection obtained through a line may appear rather low in many cases but actually it is quite sufficient; any increase does not increase the accuracy of matching, for the limit of accuracy is imposed by irregularities of the plate as will be described later. The sensitivity is fully sufficient to reach this limit.

§ 3. METHOD OF USE

The procedure for finding match points is very simple. First with the plate on the stage the necessary focusing adjustments are made; the plate is then moved so that an unexposed part, in the neighbourhood of the match point, is imaged on the slit and then the plate *G* is adjusted until the galvanometer shows no deflection. Then the plate is placed so that the image of the dividing-line between the two spectra falls on the engraved line on the slit, and the stage is traversed until the deflection of the galvanometer becomes zero; the match point is then imaged on the slit in the screen. The stage can be traversed as a whole along the guides; in the neighbourhood of a match point it can be clamped and a more sensitive adjustment can be made by means of the screw *F*, which is 25 mm. long.

The wave-length corresponding to the point at which the match point occurs must be determined. In the case of line spectra the match point will in general fall between two lines; these lines can be marked directly on the plate, and their wave-lengths can afterwards be determined by reference either to a map or to the wave-length scale on the plate. But in the case of continuous spectra a different procedure must in general be adopted; the difficulty is that, although a mark could be made on the plate to indicate the position of the match point, reference to the usual wave-length scale at the top and bottom of the plate is not sufficiently accurate owing to the curvature which is impressed on large plates in most spectrographs. The difficulty is overcome, however, by arranging that a line spectrum shall be taken with each absorption exposure, to lie in juxtaposition with the twin absorption spectra. This spectrum will appear on the screen with the absorption strip under examination. When the match point has been found the reading of the screw is observed; then the plate is traversed until one of the lines in the line spectrum falls on a continuation of the slit and the screw reading is again read. The line can be marked and later identified, and since the distance of the match point from this line is known in millimetres, the wave-length of the match point can be determined by simple interpolation or by the application of a Hartmann formula, according to the accuracy required.

§ 4. RESULTS

It was found that match points could not be determined with the precision that had been expected. In the case of line spectra it had been anticipated that in general the match point would not fall definitely on a line but would be between two lines. It had been expected that measurements on the background between the lines would not be of much value, but that some method of interpolation could be usefully applied.

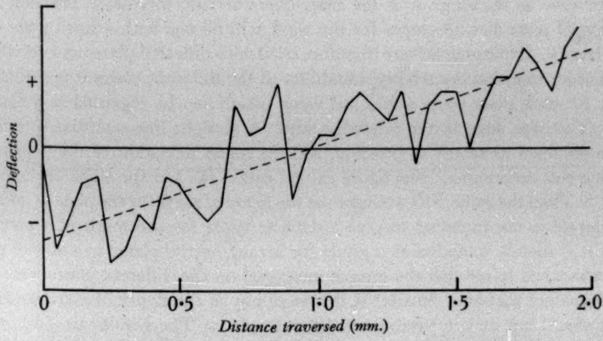


Figure 4.

It was found, however, that in cases where the absorption band was shallow, an equivocal kind of result was obtained, in which several successive lines gave deflections falling at random on either side of the zero; all that could be said in such cases was that the match point lay somewhere in the midst of this group of lines. It was thought that this might be due either to pole effects arising from the incorrect adjustment of the spark source in the photometer, or to irregularities in the photographic plate. Readings were therefore taken on plates obtained with a continuous spectrum on the Hilger-Nutting spectrophotometer, with a camera fitted in place of the usual arrangements for visual readings. Any irregularities found here would presumably be due only to irregularities in the plate.

The actual deflection observed was taken at a large number of settings in the neighbourhood of match point. The area of each spectrum measured at a setting was a rectangle measuring 0·15 × 0·05 mm. approximately, and the readings were taken at intervals of 0·05 mm. Figure 4 shows the result obtained with an Ilford auto-filter plate developed with metol-quinol. It will be observed that there is a region nearly 0·9 mm. in length in which the line crosses and re-crosses the zero line. It is possible to draw a straight line which can be regarded as making an approximate fit with these points, and the point where this line crosses the zero line can be regarded as the match point. The match point can probably be determined fairly accurately in this way, but it is obviously out of the question for this procedure to be regarded as one to be generally adopted as it would be far too laborious.

The irregular nature of the line is presumably due to graininess of the emulsion, a factor which varies with both the developer and the type of plate. It is therefore clear that a combination of plate and developer showing the minimum of grain effect should be chosen. At the same time it is clear that the greater the slope of the line, the more clear-cut will become the match point. For a given absorption spectrum the slope will depend upon the contrast given by the plate; that is, it will increase as the slope γ of the plate-characteristic increases. The best combination of plate and developer for the work will be one with a small grain effect and a large γ. Experiments were therefore tried with different plates and developers.

$\bar{D}$
S

In order to assess the relative suitability of the different plates it is desirable to derive for each plate some numerical value which can be regarded as a figure of merit. This was done in the following way. A straight line was drawn to fit the points obtained as nearly as possible, and the mean deviation of the points from this line was determined; this figure can be called $\bar{D}$. Let the slope of the line be called S. Then the value $S/\bar{D}$ was taken as the *figure of merit* for the plate in question. This figure is not intended to give a definite value for the merit of a particular plate; it is merely intended as a guide for arranging the plates in order of merit. Care was taken to see that the images measured on the different plates were all of approximately the same density; if this were not so the figure of merit as defined above would not be comparable for different plates. The results are given in the table.

Table

Plate	Developer	Figure of merit
Auto filter		0·17
Rapid process panchromatic	Potash-hydroquinone	0·34
	Glycine	0·41
	borax	0·25
Panchromatic half-tone	Potash-hydroquinone	1·0
	Glycine	1·0

There seems to be no point in trying a further selection of plates. The panchromatic half-tone plate was chosen as it had a higher value of γ than any other plate we could find; the rapid process panchromatic was also chosen on account of its comparatively high γ. Plates with a lower γ would need to have an extremely low grain effect to compete with these, and since the grain effect with the panchromatic half-tone plate seemed reasonably small it was concluded that a better plate was unlikely to be found among those with a lower γ.

The table shows that the panchromatic half-tone plates have the same figure of merit whether developed with glycine or potash-hydroquinone. Glycine is a developer which gives less grainy results than does potash-hydroquinone; but it also gives less contrast, and this disadvantage counterbalances the advantage of the reduced grain. Borax was found to have approximately the same grain as glycine but to give even less contrast.

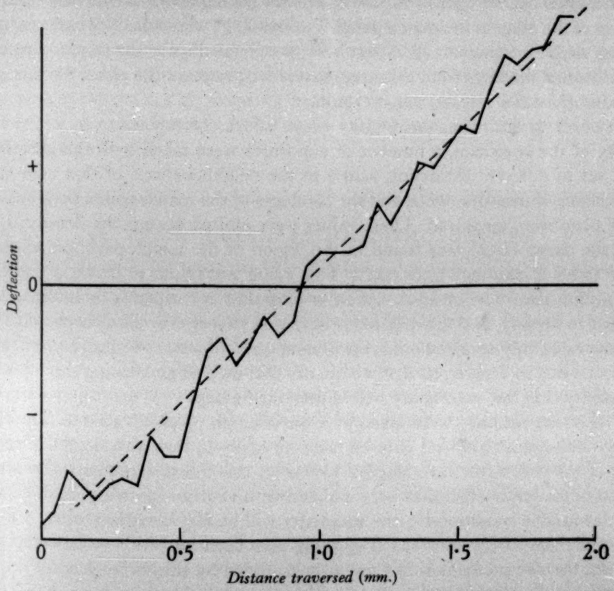


Figure 5.

We therefore conclude from these results that the best plate to use is the panchromatic half-tone; as developer potash-hydroquinone is preferable to glycine because it is more rapid in action. Figure 5 shows a plot of results obtained with this plate; it is directly comparable with figure 4, for the photometric conditions are the same and the density of the image at the match point is the same. The great superiority of the panchromatic half-tone plate is obvious.

The next stage is to determine what precision is actually obtainable in the determination of densities under ordinary conditions of use. With a coloured filter as specimen, a number of exposures were taken under the same photometric conditions and the match points were determined on the instrument. The procedure was to watch the galvanometer deflection as the screw was turned and to observe the screw-reading when the match point was considered to be attained after the small fluctuations due to grain effect had been taken into consideration. A photograph of the copper arc was taken with each continuous spectrum, so that the two spectra were closely juxtaposed; a line in the copper spectrum then served as a reference point and the distance of the match point from this line was found by turning the screw until the projected image of the line fell on the slit in the screen.

Two sets of three spectra were examined; the only difference between the two sets was in respect of time of exposure, one set having about three times the transmission of the other at the match point. Two readings were taken for each spectrum, making twelve readings in all. These results were averaged; the mean deviation of the individual readings from this average was 0·05 mm. on the plate; the maximum deviation from the average was 0·11 mm.

In order to find what uncertainty these values correspond to in terms of the density of the specimen, a number of exposures were taken with the photometer circle set to different extinction values in the neighbourhood of that used in the above series of measurements, and the distances of the match points from one edge D of the plate were measured. These values were plotted against the density D, and from the curve dD/dS was found in the region of the match point on which the above series of readings were taken. This value was found to be $0{\cdot}1/2{\cdot}5$ where S is in millimetres. The mean deviation of 0·05 thus corresponds to an uncertainty of 0·002 in density, and the maximum deviation to just over 0·004 in density. This extreme value only occurred once; the nearest one to it was 0·08 mm., corresponding to about 0·003 in density, so that we can say that on a single reading there is a high probability that the uncertainty will be less than 0·003.

The above readings were taken at a wave-length of about 4000 A. The figure 0·003 for the sensitivity can only be taken to apply to that wave-length, since the value of γ varies with wave-length. However, this figure represents a sensitivity several times greater than has been claimed with visually spotted plates under the most favourable conditions(1); the sensitivity will be less in regions where γ is less, but as the sensitivity with visual spotting also decreases with γ the comparison between the two methods is not seriously modified by this circumstance. We can, therefore, make a general claim that the photoelectric method increases the accuracy of spotting several times over.

It does not follow from this that densities can be determined photographically with an accuracy of 0·003, for it has not been proved that the photometers in use are sufficiently accurate. They have been proved to be accurate within the limits which are imposed by visual spotting, but tests made with the more sensitive photoelectric method of matching may show that there is room for improvement in them. If this should prove to be the case, it does not follow that the higher accuracy of photoelectric spotting is wasted, for, in some applications of absorption spectrophotometry, sensitivity in detecting small differences in absorption is more important than accurate determination of absolute density.

When the light-source gives a line spectrum the sensitivity of matching is not so good. The match point will in general be found to fall between two lines, which may be separated by several angstroms. The match may be made on the background if it is sufficiently intense, but in general this will not be possible because the background is usually underexposed with the result that the contrast between the spectra is low. In the absence of the grain effect a fairly accurate interpolation between the two lines could be made, when the individual densities of the images are known. But the presence of the uncertainties due to grain effect reduces the accuracy obtainable. The accuracy may be still further reduced by the presence of the pole effect in the light-source; even with the best-designed instruments small uncertainties due to this effect are likely to appear.

We therefore conclude that photoelectric matching cannot be used to the best advantage when the light-source gives a line spectrum. But when it is used with a continuous spectrum a high accuracy in observing match points can be attained. Further, the operation is not so fatiguing as is visual spotting, although it is perhaps a little slower.

§ 5. USE AS ORDINARY MICROPHOTOMETER

The instrument is easily adapted for ordinary microphotometric measurements, which greatly increases its range of usefulness. This can be carried out in two ways. A double-pole double-throw switch is connected in the circuit in such a way that when it is put over the two cells become connected in parallel, instead of in the way shown in figure 3. The galvanometer then shows the sum of the currents due to the light passing through the upper and lower parts of the slit, instead of the difference between them. Alternatively, the photocell house can be removed and replaced by one containing a single photocell mounted directly behind the slit without any optical system; this gives a rather greater sensitivity than the first method, where nearly a third of the light from the filament is lost owing to the black band painted on the rhomb behind the slit.

The selection of single lines for measurement is obviously an easy matter; it is only necessary to adjust the position of the plates, with the viewing-lens in the beam, until the image of the line to be measured is seen to fall on the slit. As the amounts of light which fall on the photocell are very small, the current read by means of the galvanometer is very closely proportional to the light-intensity even though the galvanometer-resistance is as high as 450 Ω.

The sensitivity of the system is such that with the slit at 0·1 mm. the deflection obtained through the clear part of a plate is 220 mm. Its performance is indicated by figures 6 and 7. Figure 6 illustrates its definition. The object in this case was a piece of silvered glass with parallel lines cut on the silver. The lines were ruled 0·05 mm. apart and were approximately 0·025 mm. wide, so that the object consisted of a series of opaque and clear sections, each approximately 0·025 mm. wide. Figure 7 shows the result of a run on the 3100-A. triplet in the iron arc spectrum on a plate taken on the Hilger E 1 spectrograph.

Another application of this instrument is as a photomeasuring micrometer. Suppose one wishes to measure the distance apart of two lines; it is only necessary to set the plate on the stage so that the image of one of the lines falls on the slit and then to traverse the plate until the image of the other falls on the slit, and to read off from the screw the distance traversed. The accuracy of this result is limited by the accuracy of the screw, which is correct to 1/100 mm., but this is sufficient for nearly all purposes.

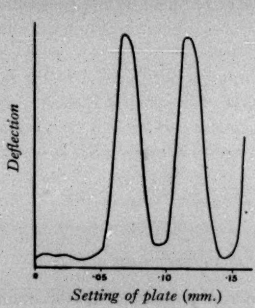


Figure 6. Engraved silvered glass plate.

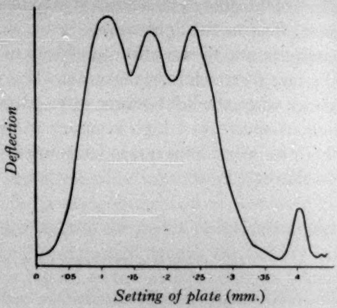


Figure 7. 3100 Fe triplet.

§ 6. ACKNOWLEDGMENTS

Acknowledgments are due to Messrs Adam Hilger Ltd., in whose laboratories this work was carried out, and at whose suggestion the results are published. I have had the benefit of useful discussions with Mr F. Twyman, F.R.S., and Dr F. Simeon, while Mr W. J. McCarthy and Mr R. A. C. Isbell have assisted ably with experimental work.

REFERENCE

(1) Twyman and Lothian. *Proc. phys. Soc.* **45**, 653 (1933).

DISCUSSION

Lord Rayleigh. I should like to ask the author why he uses colour-sensitive plates, which are in general more troublesome to manipulate? I have found that for photometric purposes plates coated on plate glass are much to be preferred, owing to the uniform thickness of the film.

Author's reply. Plates were selected by consideration of grain-size and contrast. Information concerning these points was obtained from certain data supplied by Messrs Ilford, some of which are reproduced in the paper by Twyman and Lothian referred to. The question of colour-sensitivity was not considered at all; it just happened that on the list of plates for which data were available the two which appeared most promising were panchromatic. I do not think that, for any particular type of emulsion, the grain-size and contrast would be markedly different in a panchromatic and a non-panchromatic plate, but I have no definite information on this point; in the list of plates for which data were obtained from Messrs Ilford, panchromatic and non-panchromatic plates of the same emulsion type were not included.

I have had no actual experience of the use of plates coated on plate glass. It seems probable that their use would increase the precision of absorption-measurements made by this method.

534.22

ON THE BEHAVIOUR OF SUSPENDED PARTICLES IN AIR, AND THE VELOCITY OF SOUND AT SUPERSONIC FREQUENCIES

By E. B. PEARSON, B.Sc., Assistant Lecturer in Physics,
University College, London

Communicated by Prof. E. N. da C. Andrade, August 2, 1934. Read November 16, 1934, by Prof. Andrade in the absence of the author.

ABSTRACT. In accordance with theory, particles of the size prevailing in ordinary cigarette smoke are found to act as obstacles in air vibrating at supersonic frequencies. This action leads to the formation of figures at the nodes in a resonance tube, from which measurements of the wave-length can be made. These, with a knowledge of the frequency of each piezo-electric crystal used to maintain the oscillations, enable values of the velocity of sound in air to be found at various frequencies in a range from 92·2 to 801·6 kc./sec. The values found show a definite dispersion of sound in this region, contrary to the results of previous experimenters with the Pierce acoustic interferometer. There are two maxima which are attributed to the separate effects of oxygen and nitrogen.

§ 1. INTRODUCTION

It has been shown by E. N. da C. Andrade(1) that any obstacle in vibrating air of sufficient amplitude is surrounded by a vortex system which leads to forces between two neighbouring obstacles. Two spheres whose line of centres is in the line of the vibration vector are in equilibrium at a certain distance apart: if the line of centres is transverse to the vibration vector the spheres are brought into contact. He points out that how far a particle acts as an obstacle, or is carried backwards and forwards by the medium, depends upon the size and mass of the particle and upon the frequency of vibration. For instance, cork-dust particles of the size used in his experiments take up about 96 per cent of the amplitude at 120~, and 50 per cent at 850~; but smoke particles, whose average diameter as determined by the rate of settling of a cloud was about 1μ., take up about 99·9 per cent at 2000~, and so can be used as tracing-points at this frequency. It is pointed out, however, that at a frequency of 300000 kc./sec. smoke particles of radius 1μ. should take up only 18 per cent of the motion of the air, and should thus act as obstacles. The formation of dust figures at sonic frequencies has been shown by Andrade(2) to be due to vortices around obstacles combined with a general circulation between node and antinode, which takes place in a vibrating column of gas enclosed in a tube. If this circulation takes place at supersonic frequencies smoke particles should therefore give rise to figures similar to some of the dust figures observed in a Kundt's tube. The experiments described in this paper were undertaken to investigate the behaviour of smoke particles in air vibrating at very high frequencies, with the particular object of seeing whether it was possible to obtain figures whereby the nodes or antinodes could be fixed and the velocity of sound deduced.

§ 2. DESCRIPTION OF APPARATUS

Piezo-electric quartz oscillators were used as the source of supersonic vibrations, because they maintain a very constant frequency, unaffected by temperature changes, and in addition are fairly easy to manipulate. The crystals used were supplied by Messrs Adam Hilger, and final frequency-measurements at the key frequencies were made by the National Physical Laboratory. The oscillating and sustaining circuit employed to excite the crystal is shown in figure 1; it was the standard Cady one, modified by the addition of a reaction coil. This was found necessary with the L.S. 5 valve used. The components were of the type used in standard wireless sets. To obtain adequate power a plate voltage of 360 was applied to the valve together with a negative grid bias of 20 volts, giving a plate current of 50 mA., the former being supplied by a battery of accumulators. The tuning and reaction coils were of the Igranic plug-in type and were mounted in a holder giving variable coupling, while variation of frequency was obtained by means of a Dubilier air condenser, the capacity of which was 0·0005 mF.

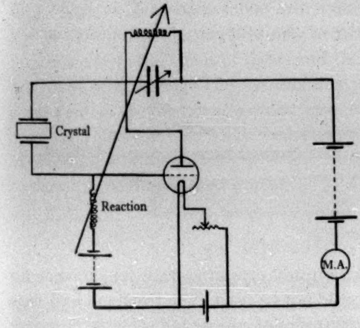


Figure 1.

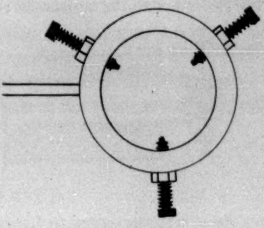


Figure 2.

All the crystals were of the same diameter, 1·9 cm., and were cut with a shallow groove in the mid plane parallel to the ends. The method of mounting is shown in figure 2. A brass ring, 2·7 cm. in internal diameter, and 4 cm. in external diameter, carries three pointed screws which grip the crystal in the groove and are secured by locking nuts. The crystals were held loosely, as otherwise they do not oscillate freely. To the brass ring is screwed a rod, held in a stand so as to allow the position of the crystal to be adjusted.

The tube containing the air to be set in vibration was of internal diameter 2 cm., and was surrounded by a water-jacket to eliminate convection. The jacket consisted of two plate-glass sides, measuring 53 cm. × 5 cm., cemented to a framework which had been cut from brass square-section tubing, and constituting the ends and base. This is shown in figure 3. The ends were brass plates pierced with circular holes, through which were soldered short lengths of brass tube marked *A*, *A* in the figure. The glass resonance tube passed through these tubes, at the points *B*, *B*, and water-tight joints were made with rubber collars of pressure tubing.

The frequency being constant for any given crystal, it is necessary to be able to vary the length of the resonant air column in order to make it contain a whole number of half-wave-lengths. At the very high frequencies, where the half-wave-length is of the order of 0·5 mm. and less, the adjustment of length is relatively unimportant, but for half-wave-lengths of 1 mm. and more it is essential. It was effected by means of a brass plunger, 2 cm. long, which fitted the tube as tightly as was compatible with easy movement and was soldered to the end of a length of 4-mm. brass tubing fitting loosely through the collar marked *C* in figure 3. A small hole was drilled through the centre of the plunger so that smoke could be introduced into the resonance space.

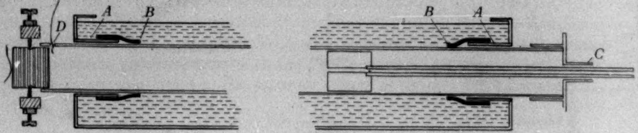


Figure 3.

As it was necessary to have an uninterrupted crystal surface for radiating the waves down the tubes, large electrodes could not be used. The front and back faces of the crystal were coppered, and small electrodes of copper foil soldered to springy wires made light contact with them. The electrode for the radiating face passed through a small hole, marked *D* in the figure, in the wall of the glass tube.

When a crystal was first tried out the correct sizes of tuning and reaction coils were found with the aid of a Sullivan heterodyne wave-meter, supplied with calibration curves for use with a series of plug-in inductance coils. From the nominal frequency of the crystal the wave-length of the electric waves in the oscillator circuit was calculated, and from the calibration curves the reading of the wave-meter appropriate to this frequency was found. The search coil connected to the wave-meter was fixed near the coils of the oscillator, whose tuning and reaction coils were changed until the heterodyne note was heard with a convenient setting of its condenser. This procedure was necessary because most crystals could be made to oscillate at several frequencies, by using sufficient reaction, but it is essential to use the fundamental mode, since that is the only one that can be maintained over a fairly wide range of capacity-change in the tuning circuit. The wave-meter was not sufficiently accurate for the purpose of the sound-velocity measurements described later, and the crystals were therefore sent to the National Physical Laboratory for accurate determination of the frequencies.

§ 3. ILLUMINATION OF THE TUBE

To render visible the movements of the smoke the image of a good optical slit, illuminated by an arc and condenser, was formed by means of a lens of focal length 50 cm. and aperture 11 cm. The flat beam in the neighbourhood of the image was used in various ways. For purely qualitative observations the whole length of the tube was illuminated through a plate glass window closing the end, this window being pierced by a hole which allowed a glass rod to pass through it. A glass disc waxed to this rod acted as the reflector of the sound waves, instead of the brass plunger shown in figure 3, and allowed the length of the resonant column of air to be altered. The beam from the arc could thus be thrown down the axis of the tube and the phenomenon could be viewed as a whole.

This method of illumination was not intense enough when the spacing of the smoke figures was to be measured or when the figures were to be photographed, and for these purposes the following arrangement was used. The tube was set up perpendicular to the beam of light, and the image of the slit was formed along the axis of the tube through the surrounding water jacket. By this means a length of about 4 cm. of the inner walls of the resonance tube could be intensely illuminated. A screen of semicircular blackened metal placed underneath, in the water, served to cut off most of the high lights from the curved walls of the tube, and also helped to show up the illuminated portion.

§ 4. PHENOMENA OBSERVED

Preliminary experiments were made at a frequency of 92·2 kc./sec. with the longitudinal illumination already described. The approximate wave-length of the sound-waves in air at the temperature of the water in the bath was calculated, and the distance from crystal face to glass reflector was adjusted so as to contain a whole number of half-wave-lengths. The small electrode, which passed through the hole in the end of the tube, was arranged to be about 3 mm. away from the end of the crystal, while the other electrode just touched its back face. This disposition of the electrode was found to give best results with the crystals whose natural frequencies were 92, 147 and 220 kc./sec.; their relatively large vibrations would be damped if there was pressure on the surfaces. With crystals of higher frequency, however, both electrodes must touch the faces.

Smoke, dried by being blown through a long calcium-chloride tube, was then introduced into the length of tube between crystal and reflector, and the supersonic vibrations were started by varying the tuning condenser about the critical position found by the wave-meter method already described. The smoke was seen to circulate violently, and in a few seconds coagulation of the particles took place at equally spaced intervals, causing the formation of rings which extended up the walls of the tube and sometimes completely round them. When the oscillations were stopped the particles ceased to coagulate, and floated in the tube at random, but as soon as the vibrations were restarted the rings formed again. When the crystal was kept oscillating for about half a minute a slight deposit was formed on the walls of the tube at the positions of the rings, and this remained after the oscillator had been switched off and the surplus smoke had been blown out of the tube.

Photographs were taken of the rings when they were in the process of formation, and also of the traces left on the walls of the tube. The earlier attempts were made with a Beck attachment fitted to the eyepiece of a microscope giving a magnification of about 6, by the method of longitudinal illumination, but insufficient light was reflected from the particles to give any record with exposures of 4 or 5 seconds, even when the fastest plates, 2000 H and D, were used. This length of exposure was the maximum possible when the rings were in process of formation, for they soon became indistinct owing to the deposition on the tube walls.

The transverse illumination, using slit and lens, caused too many high lights due to reflections from the outside of the tube, and consequently a square-section tube was made from four pieces of photographic negative glass, each 16 cm. long, cemented together at the edges to give an internal side of 2 cm. One end was closed with a brass plate, provided with a hole to allow a glass tube waxed to a square glass reflector to pass through it. By this means the length of resonant air column was varied. The other end was closed with a piece of card, with a round hole cut in it to allow the crystal to project inside. Sufficient light was obtained by using two identical optical systems of the type described, both being horizontal and making equal small angles with the normal to the tube. By this means photographs of the smoke figures seen during the oscillation of the air were taken with an exposure of 3 seconds. The result is shown in figure 4, where the line of the reflector forming the closed end is marked by an arrow. The curvature of the smoke planes was due to the strong circulation, which disturbs their alignment, but is only seen in the square tube. In a circular tube the planes are all quite flat and perpendicular to the axis. In all probability secondary disturbances in the air, due to the edges of the square tube, are responsible for the appearance of the planes in figure 4.

Examination of the traces left on the walls of the tube after oscillations had been stopped and surplus smoke had been blown out shows that they mark the nodes. Figure 5 is a photograph of these traces, formed in the circular tube at a frequency of 92·2 kc./sec. The picture was taken with the Beck attachment on a microscope, illumination being with the arc, slit and lens. Most of the high lights have been cut off by the blackened metal screen placed under the tube, but a few remain and are marked as such. This arrangement allows but a small width of the wall of the tube as being visible, but actually the traces extend completely round it. The sharp line marked by an arrow is the end of the brass reflector, and hence represents a node, so it is clear from measurements of the spacing of the traces that they themselves are at nodes. The double line forming each trace can be taken to represent the heaps of dust to either side of a node which, in a Kundt's dust tube, form the figures described by Andrade as "eyes". The double lines are shown more sharply in figure 6, which was taken at an early stage in the formation of the traces. As the amount of deposit increases the lines broaden so as to overlap, and at frequencies above 400 kc./sec. they are so close that the double structure is

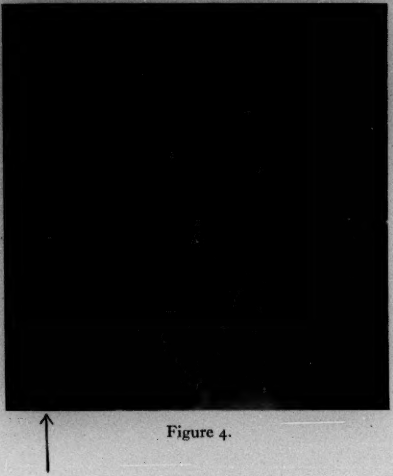


Figure 4.

High light due to reflection

Figure 5.

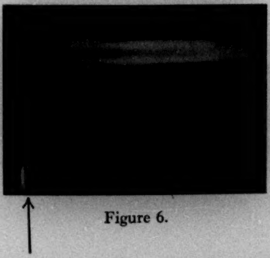


Figure 6.

rarely seen. Before a fresh set of photographs was taken the tube was cleaned out by means of a wad of cotton wool moistened with methylated spirit, and then dried with a soft cloth, pushed in on the end of a rod. The crystal was then replaced in position, and fresh smoke was blown in.

The phenomena here described can be explained on the view that smoke particles should act as obstacles in a gas vibrating at a very high frequency, and behave as cork particles do at sonic frequencies, in the manner pointed out by Andrade(1). Two particles which act as obstacles in a vibrating medium are brought and held together with their line of centres transverse to the vibration vector. It appears from the work of Whytlaw-Gray* that smoke-particles which come into contact as a result of ordinary casual impacts adhere to one another, so that it is to be expected that particles brought together by the forces generated by the vortex motion in the medium will likewise adhere. The effect will be more marked with the larger particles formed by the adhering of two of the original smoke particles, so the effect is cumulative, and rapid coagulation will take place to form the smoke planes seen when the supersonic vibrations are in progress. The fact that the planes disappear when oscillations cease shows that the coagulation is not of the ordinary type which is observed in a smoke cloud when this is left for a sufficient length of time; the coagulation must be due to minute vortices. When the oscillations are allowed to persist, the cumulative coagulation results in depositions, analogous to the eyes seen in a Kundt's tube, on the walls of the tube on either side of the nodes.

The phenomenon cannot be followed in all its details, as it was by Andrade in his work on dust figures formed at sonic frequencies, for with a crystal neither frequency nor intensity can be varied, and both of these factors enter into the mechanism of the formation of the various groupings and arrangements of particles in a Kundt's tube(2).

§ 5. DETERMINATION OF SOUND-VELOCITY

The distance between the mid lines of successive eyes deposited on the walls of
the tube represents a half-wave-length $\frac{1}{2}\lambda_t$ of the sound in air, at the particular λ_t
frequency of the crystal. This being accurately measured and the frequency being
known from the report by the National Physical Laboratory on the crystal, a value
of the velocity V_t at the temperature t° C. of the water-bath can be found from the V_t, t
relation

$$V_t = n\lambda_t,$$

where n is the frequency. n

The positions of forty or fifty nodes were noted by means of a Cambridge travelling microscope giving a magnification of about × 6 and having graduations down to 0·01 mm. on a rotating head attached to the micrometer screw. Estimations to 0·002 mm. could easily be made, and readings were always taken from the closed end of the tube towards the crystal, but not nearer to the crystal than 6 cm. so that the region where the sound-velocity has been found by previous experimenters to increase slightly on account of proximity to the source was avoided.

---

* See e.g. Whytlaw-Gray and Patterson, *Smoke*.

V_0

The set of readings of nodal positions was divided into two equal groups, in the usual manner, and thus an accurate value of $\frac{1}{2}\lambda_t$ was found. The velocity at $t°$ C. was then evaluated and the velocity V_0 at 0° C. was deduced by means of the relation

$$V_0 = V_t - 0{\cdot}61\,t,$$

where V_0 and V_t are in metres per second, and t is in degrees centigrade.

At sonic frequencies the diameter of the tube influences the velocity, and the Kirchhoff correction must be applied in order to get the value of velocity in free space. This correction is expressed in the formula

$$V_T = V_A\{1 - 0{\cdot}54/2r\sqrt{(\pi n)}\},$$

V_T, V_A, r

where V_T and V_A are the velocities in tube and free air respectively, and r is the radius of the tube in cm. In a tube of radius 1 cm., with a frequency of 92·2 kc./sec., which was the lowest used, the correction works out to be 1 in 2000, while at higher frequencies it becomes rapidly less. As the probable error in a series of values of V_0 came to about 1 in 800, the correction was neglected.

At the first frequency used, 92·2 kc./sec., a mean of eight values of V_0 gave a probable error of 0·4 m./sec. It was decided to measure the velocities at higher frequencies, and for this purpose eight more crystals, covering a range 127 to 1020 kc./sec., were employed. On repetition of the procedure already described it was found that similar phenomena were observed at each higher frequency, traces being left on the walls of the tube at the nodes, and these were measured with the microscope to evaluate $\frac{1}{2}\lambda_t$. Most of the crystals oscillated quite readily, and in some cases so violently that the crystal broke. Above 500 kc./sec., however, it was often a somewhat difficult matter to get oscillations sufficiently powerful to form the smoke figures, and careful variations of reaction coupling, filament current, and pressure of the electrodes on the crystal faces, were all necessary in order to bring about the desired result, except in the case of the 1020-kc./sec. crystal. In this instance no figures could be detected, although there was a fairly well-marked circulation in the smoke.

The value of V_0 obtained at each of the other frequencies was taken as the mean of eight determinations. A typical example of such a set is given in table 1.

Table 1

Frequency $n = 310{\cdot}198$ kc./sec.

Temperature (° C.)	Wave-length (mm.)	V_t (m./sec.)	V_0 (m./sec.)
16	1·104	342·5	332·8
16	1·106	343·1	333·2
16	1·103	342·1	332·4
16	1·104	342·5	332·8
17·5	1·108	343·7	333·0
18	1·112	344·9	333·8
15	1·104	342·5	333·3
17·5	1·109	344·0	333·3

Mean $V_0 = 333{\cdot}1 \pm 0{\cdot}2$ m./sec.

The values of V_0 obtained at each of the eight frequencies are given in table 2.

Table 2

n (kc./sec.)	V_0 (m./sec.)
92·20	330·7 ± 0·4
127·53	331·1 ± 0·4
219·28	332·6 ± 0·7
310·19	333·1 ± 0·2
356·78	331·9 ± 0·5
485·34	330·4 ± 0·5
628·96	331·1 ± 0·6
801·67	330·8 ± 0·6

The result of plotting V_0 against log n, which gives a better-spaced scale than n itself, is shown in figure 7.

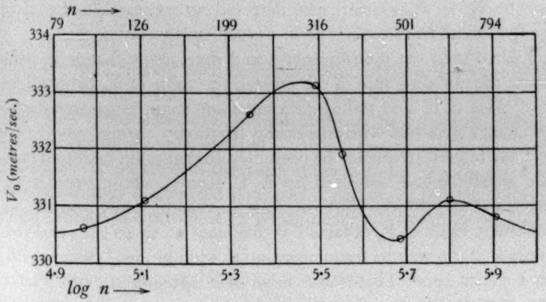


Figure 7.

§ 6. DISCUSSION OF RESULTS

The first measurements of V_0 were made at 92·2 and 310·2 kc./sec. and an increase above the normal value was found at 310·2. As air is a mixture, it was thought that each of the two principal gases might show its own maximum value of V_0 at an appropriate frequency, and as no further maximum appeared up to 485·3 kc./sec. on using the intermediate frequencies, the values above it were used, when the second small maximum was found at 630 kc./sec. The supposition that this is due to oxygen while the higher peak at 290 kc./sec. is due to nitrogen is further strengthened by the fact that the peak at 290 kc./sec. is almost exactly four times as large as that at 630 kc./sec.

A number of experimenters have worked on the problem of velocity-changes due to dispersion at supersonic frequencies, and although no recorded results show a marked change in air, it must be stressed that the values of V_0 at any given frequency, as found by different observers, differ appreciably amongst themselves.

With the exception of C. B. Vance(3), who used lycopodium to indicate the nodes, all other experimenters have employed the Pierce acoustic interferometer, in which the waves emitted from the oscillating crystal are reflected back on to it from a plane set parallel to it and capable of being moved perpendicular to its face by means of a micrometer screw. The difference in phase between reflected and emitted waves, which depends on the position of the reflector, causes the plate current of the oscillator to pass through a series of maxima, the distance moved through by the reflector between successive maxima being equal to one half-wave-length of the sound waves. The first important measurements of velocity in air were made by G. W. Pierce(4), who covered the range 90 to 1040 kc./sec. with four crystals having frequencies of 98·2, 205·6, 610·2 and 1034 kc./sec. At 610·2 kc./sec. appeared a very slight maximum value of V_0, 331·8 m./sec., to which Pierce himself attached no importance, since he considered it to be within the range of experimental error. His experiments were repeated in air(5) and later in oxygen(6) by W. H. Pielemeier, who detected no appreciable change of velocity in air at 303, 389 and 655 kc./sec., or in oxygen up to 316 kc./sec.

C. D. Reid(7) used the interferometer and detected a change in velocity with distance from the radiating face of the crystal, the value of V_0 being slightly higher when found close to it, and falling to a constant value at greater distances. The amount of change gets less with increasing frequency, being only 0·15 m./sec. at 140 kc./sec., the highest frequency he used. Up to that point he found no dispersion in air. The interferometer was used by P. T. Kao(8), who covered a range 40 to 1000 kc./sec., and also noticed slight increase in velocity quite near the source. He took the constant value of V_0 obtained at distances 30 to 40 cm., and this value of V_0 was constant at the various frequencies up to 1000 kc./sec., being 331·85 m./sec., correct to 0·1 per cent. This value is in fair agreement with Pierce's above 100 kc./sec., but the latter's values below this frequency are in excess of it.

The experiments of Vance(3) up to 200 kc./sec. gave a constant value of V_0, but his method seems open to criticism. Lycopodium powder was dusted into a tube fitted with a plunger in order to vary the length, and the open end was held close to the radiating face of the crystal oscillator. The powder was said to collect at the nodes and their spacing was then measured with a travelling microscope. Vance says that the nodes were so indistinctly marked that too great an error came in if the position of each was noted, and consequently only every tenth could be taken. This fact seems to indicate that the method is not adapted to accurate determinations of wave-length, and hence the figures for velocity cannot be very reliable. In any case, as the highest frequency used was only 200 kc./sec., the region of dispersion as found by the present writer had not been reached.

With carbon dioxide definite evidence of supersonic dispersion has been found. Pierce(4) obtained a rise in velocity of about 0·5 per cent between 98 and 205 kc./sec., but the work was incomplete and no measurements were made between 205 and 1034 kc./sec., at which latter frequency the gas was opaque to the sound and hence no figure for velocity was recorded. H. O. Kneser(9) used the interferometer method with carbon dioxide between 60 and 1480 kc./sec. His value for V_0 in the region of 100 kc./sec. agrees fairly closely with that of Pierce, but above it he found a rapid increase of magnitude far larger than that detected by Pierce and reaching 4 per cent at 350 kc./sec. Above this frequency the velocity remained constant and did not fall to the normal value.

The results shown in the present paper thus differ very widely from those obtained for air in the past with the interferometer. The slightly higher values of V_0 that have been observed near the source cannot account for the increase found by the author at 290 kc./sec., and in any case the measurements of the spacing of the smoke-traces were never made near the crystal. Both the maximum at 290 kc./sec. and the minimum at 480 kc./sec. represent changes well above the experimental error, but no theoretical explanation can be offered to account for them. However, it seems a significant fact that these results, which are the first to show marked dispersion in air at frequencies above 100 kc./sec., have been obtained by a method which is the most direct of those so far used, depending as it does on a means of marking nodes with a material substance, exactly as has been done at sonic frequencies with cork dust. In addition, the length of the resonant air column is kept approximately constant during each determination, while in the case of the interferometer it is continually varied. As the supersonic waves attenuate fairly rapidly in air it is clear that as the reflector is moved closer to the course the reflected beam becomes more powerful and hence exerts a varying reaction on the crystal, so that it may well be that the maxima in the plate current do not always occur when the reflector moves through a distance equal to a half-wave-length, for the factors governing this change of current with phase-difference must be extremely complicated.

§ 7. ACKNOWLEDGMENT

Finally I wish to offer may sincerest thanks to Prof. Andrade, at whose instigation the problem was attempted, for his continual interest and invaluable suggestions during the course of the work, and also for the experimental facilities placed at my disposal.

REFERENCES

(1) *Proc. roy. Soc.* A, **134**, 445 (1931).
(2) *Philos. Trans.* A, **230**, 413 (1932).
(3) *Phys. Rev.* **39**, 737 (1932).
(4) *Proc. Amer. Acad. Arts Sci.* **60**, 284 (1925).
(5) *Phys. Rev.* **34**, 1184 (1929).
(6) *Phys. Rev.* **35**, 1005 (1930).
(7) *Phys. Rev.* **35**, 814 (1930).
(8) *C.R.* **193**, 21 (1931).
(9) *Ann. Phys.*, Lpz., **11**, 777 (1931).

DISCUSSION

Dr A. B. Wood. I am particularly interested in the latter portion of the paper, in which reference is made to the remarkable changes in the velocity of sound in air which occur at high frequencies. Like the author, I find it difficult to supply an adequate theoretical explanation of the results. In such a case it is important to eliminate the possibilities of a defect in the experimental method. No doubt the author has already considered the possibility of gaseous impurities, such as water vapour or CO_2, being introduced into the tube with the smoke.

The statement on page 142 of the paper that some crystals vibrated so violently that they broke whilst others were difficult to excite into vibration of sufficient amplitudes, arouses the suspicion that at any rate some of the observed wave-velocity variation may be due to the differences in velocity-amplitude of the various crystals used in the measurements. It is a simple matter to show that the velocity-amplitude $\dot{\xi}$ at the antinodal faces of a half-wave-length resonator is given by

$$\dot{\xi}=\frac{P}{E}C,$$

where P is the stress at the nodal plane, E the elastic modulus and C the velocity of sound in the oscillator. The upper limit of P is set by the breakdown stress of the oscillator. Tables give various values for quartz, an average being $0\cdot5\times10^{10}$ dyne/cm.² If $C=5\cdot5\times10^5$ cm./sec. and $E=8\times10^{11}$ dyne/cm.², this gives the following value for the maximum velocity-amplitude of the face of the crystal:

$$\dot{\xi}_{\text{max.}}=3500 \text{ cm./sec.}$$

That is, the quartz crystals used in the experiments might have any velocity-amplitude up to 35 m./sec., resulting in a pressure-amplitude of 0·14 of an atmosphere in the air near the face of the quartz. To this must be added the possibility of a further increase of velocity-amplitude due to resonance in the air column contained in the tube. Such large-amplitude particle velocities might be expected to produce serious distortion of wave-form and increased wave velocity.

In a resonant air column of the kind used in the experiments it is unlikely that much decrease of velocity with distance from the crystal-face would be observed. The antinodes in the air column will, in a reasonable length of tube, have approximately the same velocity-amplitude, unless the gas produces strong attenuation. The author's remark on page 145 that "measurements of the spacing of the smoke-traces were never made near the crystal" does not, therefore, eliminate the possibility of high wave-velocities, due to resonance of the crystal and air-column.

It appears to be desirable that further experiments should be made with the 310-kc. crystal, which gives an abnormally high velocity, excited at different voltages or used in different gases. Such experiments would reveal how much of the observed change in sound-velocity is due to variation in velocity-amplitude of the vibration and how much to the nature of the gas.

I do not wish to suggest that the observed velocity-changes at high frequencies are not due to the causes mentioned in the paper, but I think that other possibilities such as those I have mentioned should be eliminated before a satisfactory theory can be established.

Prof. E. N. da C. **Andrade**. I am much impressed by Dr Wood's calculations, which are possibly not, however, as conclusive as he believes. The crystals which break do not crack across the nodal planes but at one edge, with cracks that appear to be rather parallel to than normal to the axis of the cylinder. I am inclined, from examination of the crystals, to think that the fracture is largely due to marked weakening consequent on the cutting of the groove for holding the crystal. The effect of any minute sharp edges left on such an entrant cut is well known. Incidentally, the value given for the tensile strength of quartz in the *International Critical Tables**, quoted as "maximum observed", is about one-fifth† of that quoted by Dr Wood, and is presumably for a static load, and it further seems likely that, as only a few crystals failed, they were rather of minimum than of maximum strength. Lastly, I am in some little difficulty to see how the effect contemplated by Dr Wood could be reasonably expected to lead to a rise to a maximum followed by a drop, in the consistent way shown by all results.

As regards Kneser's theory, while, as I pointed out in reading the paper in the author's unavoidable absence, it is undoubtedly attractive and has had good success in some directions, Kneser's experiments themselves do not furnish any evidence of a maintenance of the high velocity at higher frequencies, but merely show that it is just attained. A subsequent drop is quite consistent with his experimental figures. There is, further, a difficulty on the question of absorption. Kneser himself alludes to Pierce's finding that CO_2 is practically opaque to sound waves in the region of 2×10^5 c./sec. and says that the differences of absorption are greater than is to be anticipated from the theory. He adds that the experimental result speaks rather for the occurrence of selective absorption in the region of dispersion, of which the theory can give no account. I also note that Sherratt and Griffiths, in their recent paper, while saying that they do not attach much weight to the discrepancy, nevertheless point out that with the gas, carbon monoxide, used by them "no agreement was found between the resulting sound-absorption in the gas and that to be expected from Kneser's theory". I do not think that the theoretical position is sufficiently well established to throw doubt on the experimental results of the author.

Mr J. H. Awbery. In presenting the paper, Prof. Andrade referred to the work of Kneser, whose theoretical curve of velocity as a function of frequency does not show a maximum, but rises gradually from one constant value to another. He pointed out that Kneser's observations in CO_2 did not extend far enough to verify the later part of the curve, i.e. to show that the course of the curve after the rise is horizontal. It may be of interest to note that in a paper read to the Royal Society on November 15, Dr Griffiths and Mr Sherratt provide indirect evidence of the truth of Kneser's theory. They worked at two frequencies only, but by using the theoretical formula for the curve given by Kneser, they deduce the true velocity of sound, and hence the ratio of the specific heats of the gas that they used (carbon monoxide). Their results up to 1800° C. agree with those deduced for specific heats from spectroscopic data, thus removing a discrepancy which had for some time been troublesome. This result at least makes it probable that the theory given by Kneser has in fact traced the true cause of the discrepancy, and that his theoretical curve would therefore be of the correct form.

---

* *International Critical Tables*, **4**, 22.

† Viz. $0 \cdot 91 \times 10^9$ dyne/cm. Voigt gives $1 \cdot 26 \times 10^9$ dyne/cm.

I wonder if any tests were made to ascertain whether the frequency of the crystal when coupled to the resonating air column is the same as when it was calibrated—i.e. presumably with a very different acoustic load. Even a slight alteration in frequency would suffice to affect the shape of figure 7 materially.

Mr G. G. SHERRATT. Figure 7 appears to me to be remarkable in that there is a maximum in the (velocity, frequency) curve. Since the work of Kneser and the consequent explanation of the discrepancy that has existed for a long time between spectroscopic specific heats and those determined from sound-velocity data, one would not be surprised to see an increase of velocity with frequency. But that velocity should subsequently decrease as the frequency is raised demands a further explanation. Kneser's work was founded on experimental evidence and has since received much experimental and theoretical support. Furthermore, it provides a reasonable physical picture to account for the increase of velocity with frequency.

If there is an increase over the frequency-range covered by the present experiments, it must presumably be due to the apparent disappearance of part of the rotational heat-capacity of the oxygen and nitrogen molecules. The small vibrational specific heat possessed by oxygen at room-temperatures disappears at much lower frequencies.

With regard to the query at the end of the paper as to the accuracy of the acoustic interferometer, it may be as well to state that the interferometer yields accurate half-wave-lengths if two precautions are observed. The first is that the sound-intensity must be sufficiently small for the limiting velocity to be obtained. The second is that the frequency must be independent of reflector-position.

In this connection it would be of interest to know whether the author took the precaution of verifying that the frequencies of the crystals were unaffected by variations in acoustic load.

AUTHOR's reply. In reply to Dr Wood: the only crystals that broke were those of the lower frequencies, before the velocity increased very much, which seems to indicate that the maximum amplitudes, which might be responsible for increased wave velocity, occurred at frequencies that gave almost normal velocity-values. As to gaseous impurities—the water vapour was removed, while any CO_2 would be expected to prevent a peak in the curve, for Kneser's experiments showed that the high velocity-values persisted as the frequency increased, in the case of that gas. Further experiments with different gases are already being considered.

In reply to Messrs Awbery and Sherratt: no tests were made to ascertain whether acoustic load influenced frequency, but I suggest that the frequency of a quartz crystal, which depends solely on its physical constants, would not be likely to be altered by the back pressure of the resonant air column. Kneser's figures for pure CO_2 will be checked in due course, when the apparatus for using gases other than air is built.

534.22

THE VELOCITY OF SOUND IN SHEET MATERIALS

By A. B. WOOD, D.Sc., F.Inst.P.,

AND

F. D. SMITH, D.Sc., A.M.I.E.E.,
Admiralty Research Laboratory, Teddington

Received March 16, 1934. *Read October* 19, 1934.

ABSTRACT. Chladni's figures have been produced at frequencies up to 20,000 by means of magnetostrictively excited tubes or rods of nickel. The velocity of sound in sheets of brass, iron, lead, ebonite, celluloid, cardboard and other materials has been measured with considerable accuracy by this means.

In a recent issue of *Nature** a note by R. C. Colwell refers to the excitation of Chladni's figures at high frequencies by means of a nickel rod excited magnetostrictively and arranged to set circular and square brass plates in vibration. The paper concludes with a statement that the plates "cannot take up such a high vibration as 15,000 per second."

As a matter of fact this method of exciting sheets of various materials into transverse vibration has been in use in this laboratory for a considerable time, particularly in connection with magnetostriction depth-sounding apparatus. The results obtained prove not only that thin sheets of metal or almost any solid material can vibrate transversely at these high frequencies, but also that measurements of the spacing of the nodes yield a tolerably good value for the velocity of sound appropriate to the thin metal sheet.

If a horizontal sheet of the material is sprinkled with sand and touched at a suitable spot with a tube or rod in vigorous high-frequency vibration, transverse waves travel over the sheet and are reflected back on reaching the edge or edges. The sheet is assumed to be large compared with a wave-length of the waves transversing it. The edge-reflected waves interfere with the direct waves and produce a stationary wave pattern in the sand. A resonating nickel tube has been found to give good results, eddy-current losses at these high frequencies being much less in a thin-walled tube than in a solid rod. With metallic sheets such as brass the stationary wave pattern is often complex, owing to interference of the direct wave with the waves reflected from the four edges of the sheet.

With ebonite, celluloid and other non-metallic sheets, however, the attenuation of the transverse wave is sufficient to leave only the simple pattern produced by the direct wave and that reflected from the nearest edge of the sheet; the waves reflected from the remoter edges are generally of insufficient amplitude to disturb the sand.

---

* *Nature*, Lond., **133**, 258 (1934)

If the sheet is circular, and is touched by the nickel tube at the centre, a series of equi-spaced sand-rings are produced from which the wave-length of the transverse vibration is easily measured. In all cases the sheet must be supported as lightly as possible so as to introduce a minimum of interference with the free vibration.

Thin nickel tubes have been used at resonance frequencies of 20,000 or more c./sec. with success, but it is advisable to select the frequency and the thickness of the sheet in accordance with the conditions mentioned below.

If the point of excitation is suitably chosen it is possible to obtain nodal lines running parallel to one edge of the plate, the measurement of the half-wave-length (distance apart of the nodal lines in the sand) being made in a direction normal to the free edge which reflects the waves. In such a case we may, with a fair degree of accuracy, write*

$$C=\frac{(1+k^2\kappa^2)^{\frac{1}{2}}}{k\kappa}\,C_t \qquad \ldots\ldots(1),$$

C_t, C
k, λ, κ

where C_t is the velocity of transverse waves, C is the velocity of sound appropriate to the sheet, $k=2\pi/\lambda$, where λ is the transverse wave-length, and κ is the radius of gyration of the cross-section of the sheet about the neutral axis of bending. When $k\kappa$ is small this reduces approximately to

$$C=C_t/k\kappa \qquad \ldots\ldots(2).$$

The accuracy of the velocity determined by this method is dependent ultimately on the accuracy with which the thickness t of the sheet and the transverse wave-length can be measured, since $\kappa=t/\sqrt{12}$ and $k=2\pi/\lambda$.

Photographs of sand figures on sheets of brass, celluloid and ebonite are shown in figures 1, 2 and 3. The relevant data for these are given in the table at nos. 5, 16 and 14. Even a relatively soft metal like lead gives very clear nodal lines, and very thin metal sheets, such as 0·004-in. stalloy (no. 7 in the table), give good results.

The method is shown at its best when large sheets of relatively non-resonant materials such as celluloid, cardboard or ebonite are used. In these cases the simple pattern obtained by reflection from the edge nearest to the source agrees remarkably well with the calculated distribution of the nodal lines. The latter are produced by the interference between the direct waves from the source and the reflected waves from the image source, and consist of a family of hyperbolae. An example of this is illustrated in the photograph (figure 3). The principal hyperbolic nodal lines are clearly defined, whilst a second family of hyperbolic nodal lines, due to reflection from the remoter opposite edge (not included in the photograph), are faintly shown crossing the principal series.

The table gives the values of the velocity of sound in various sheet materials determined by this method.

---

* See Lamb, *Sound*, equation (12), p. 123 (1910).

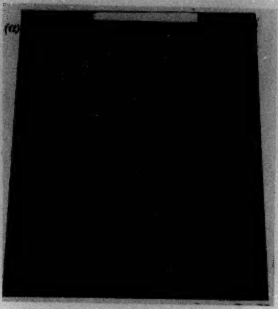


Figure 1. Brass 18 in. square and $\frac{1}{8}$ in. thick.

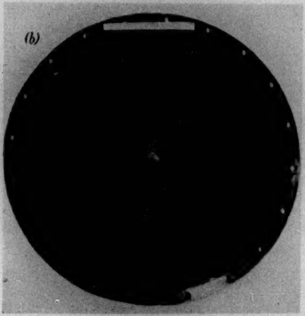


Figure 2. Celluloid 20 in. in diameter and $\frac{1}{16}$ in. thick.

(c)
Reflecting edge

Figure 3. Ebonite 3 ft. × 18$\frac{1}{2}$ in. × $\frac{1}{8}$ in.

[To face p. 150

Table 1

No.	Sheet material	Thickness (in.)	No. of half-wave-lengths counted	Total length (cm.)	λ (cm.)	N (c./sec.)	C_t measured (cm./sec.)	$k\kappa$	C (cm./sec.)
1	Brass	0·019	42	40·3	1·92	9,070	$17{\cdot}4\times 10^3$	0·0455	$3{\cdot}82\times 10^5$
2	,,	0·019	36	30	1·67	12,100	20·2	0·0521	3·88
3	,,	0·019	40	27	1·35	18,400	24·9	0·0645	3·86
4	,,	0·128	10	25	5·0	9,140	45·6	0·118	3·88
5	,,	0·128	16	27·5	3·56	18,300	65·2	0·165	4·0
6	Iron	0·095	22	55	5·0	9,500	47·5	0·087	5·46
7	Stalloy	0·004	10	5·4	1·08	9,070	9·8	0·017	5·8
8	Aluminium	0·067	12	26·5	4·42	9,500	42·0	0·0695	6·05
9	,,	0·103	10	27·5	5·5	9,500	52·2	0·086	6·1
10	Nickel	0·125	10	29·4	5·88	9,500	55·8	0·0978	5·71
11	Alpax (cast)*	0·166	12	28·5	4·75	18,260	86·6	0·160	5·5
12	Ebonite	0·046	12	8·6	1·43	18,290	26·2	0·147	1·73
13	,,	0·046	9	9·1	2·02	9,070	18·3	0·104	1·77
14	,,	0·124	10	11·5	2·3	18,200	41·9	0·249	1·73
15	,,	0·124	18	30·0	3·33	9,070	30·2	0·172	1·78
16	Celluloid (clear)	0·062	12	14·1	2·35	9,070	21·5	0·121	1·79
17	Lorival†	0·127	12	12·8	2·1	18,200	38·3	0·278	1·44
18	,,	0·260	13	28	4·3	9,070	38·9	0·277	1·45
19	Cardboard	0·070	8	10·6	2·65	9,070	24·0	0·121	1·98
20	Lead	0·010	12	5·4	0·9	9,070	8·16	0·051	1·59
21	,,	0·040	10	8·9	1·78	9,070	16·1	0·103	1·56
22	,,	0·060	10	10·7	2·14	9,070	19·4	0·129	1·51
23	,,	0·070	8	9·55	2·4	9,070	21·8	0·134	1·63
24	Keramot†	0·030	12	6·8	1·13	18,250	20·7	0·122	1·7

* Al-Si alloy. † Commercial insulating materials.

It will generally be found, where comparisons are possible, that the values of the velocities of sound in the sheet materials tabulated above are intermediate between the values obtained for the longitudinal velocities in rods and in bulk. The velocities in rod, sheet and bulk are the uni-, bi- and tri-dimensional cases in which the velocities are given by

$$\sqrt{\frac{E}{\rho}},\quad \sqrt{\frac{E}{\rho\,(1-\sigma^2)}},\quad\text{and}\quad \sqrt{\frac{E\,(1-\sigma)}{\rho\,(1-\sigma-2\sigma^2)}}$$

respectively, where ρ is the density, E Young's modulus, and σ Poisson's ratio.

DISCUSSION

Dr E. J. Irons. The measurement of velocities in sheet materials is a subject of some interest to me as it has, during the past session, engaged the attention of Mr Huffington and myself at East London College. Our aim has been to determine the velocity of a pulse produced in a material by displacements at various angles to its plane, and our purpose that of developing a method to provide constants that we hope may ultimately prove of service in moving-coil loud-speaker design. We have employed a direct timing method and hope to publish in due course an account of these and other experiments made with our apparatus.

I suggest that a reference to the source of equation (1) would render the reading of the paper easier.

Dr N. W. McLachlan. Have the authors tried their method with paper from 8 to 10 mils thick? The velocity of sound in cardboard given in the table is of the same order as the value I obtained for paper some years ago by calculation from direct measurements of elasticity and density*.

AUTHORS' reply. We shall look forward with interest to seeing Dr Irons's results. The reference for equation (1) has now been inserted in the paper. In reply to Dr McLachlan: the method as applied to thin cardboard is equally applicable to thin paper. Even tissue paper 5×10^{-4} in. thick gives clear nodal lines from which the velocity may be calculated. As we pointed out in the paper, however, the accuracy ultimately depends on the accuracy with which the thickness of the sheet is measured.

---

* *Phil. Mag.* **13**, 115 (1932).

534.323

THE MEASUREMENT OF LOUDNESS

By N. R. CAMPBELL, Sc.D., F.Inst.P.,
AND
G. C. MARRIS, B.Sc., M.I.E.E.
A communication from the staff of the Research Laboratories
of The General Electric Company Limited, Wembley.

ABSTRACT. No new scale of loudness is proposed, nor are those already proposed examined completely. The object of the paper is to examine proposed scales from one point of view only, and to determine only whether any one of these is distinguished from the remainder by being founded more firmly on facts and being more free from arbitrary convention. We are not concerned primarily with such questions as convenience or practicable accuracy.

In § 1 the facts (*a*)–(*f*) implied by all adequate scales of loudness are stated in a manner that does not (like most statements) imply that any scale is possible. It is concluded that these facts are not sufficient to distinguish from each other the infinite number of scales that might be or have been proposed. If one scale is to be distinguished from another by facts, some other facts must be introduced which are not so intimately involved in the meaning of loudness. There is an important difference in this respect between scales for tones of the same frequency and scales for all sounds.

In § 2 proposed scales for a single frequency are considered under the following headings: (i) mental estimates; (ii) equal relations; (iii) just-perceptible differences; (iv) thresholds.

The presumed facts (*g*)–(*k*) are involved in these scales and in part determine them; but the facts as known at present are not sufficient to force on us the selection of one particular scale, though they may force us to reject some.

In § 3 the problem of sounds in general is considered. Here again it is concluded that the facts, though they may limit further the choice of scales, do not determine one scale uniquely. On the other hand, if a scale for a single frequency could be agreed upon, work on the lines of Fletcher and Munson's latest publication would solve the problems peculiar to complex tones so far as it is soluble at all.

In § 4 some more general questions are treated briefly. It is urged that the arbitrary element in all scales may be advantageous as well as disadvantageous, for it permits the use of different scales for different purposes. To the question whether a scale uniquely determined by facts can ever be set up, we reply with a guarded negative; such a scale is conceivable, but all evidence available at present indicates that it is no more than conceivable.

---

§ 1. THE FACTS IMPLIED BY SCALES OF LOUDNESS

The following discussion is as little abstruse and as closely confined to the immediate question as possible. But some reference to the general theory of measurement cannot be avoided. In our use of terms (such as *A* and *B* magnitudes, regular and irregular errors, constants, accuracy and sensitivity) not defined in the text, but not unambiguous in current usage, we have followed Campbell's *Measurement and Calculation*, to which reference may be made.

The facts of audition are sometimes stated in such a way as to imply that an entirely satisfactory scale of loudness is already available. Since our object is to enquire whether such a scale is possible, this method of statement is likely to introduce confusion. We will start therefore by re-stating certain familiar facts in a way free from objection for our purpose. These facts are as follows:

Fact (*a*) (*a*) If two sounds are sounded alternately at suitable intervals, any observer who can hear both can judge whether they are equally loud or, if not, which is the louder. Such judgments define what is meant by loudness; no statement about loudness in what follows has any meaning except in reference to these judgments. In virtue of them the observer can arrange in an order of loudness all sounds which can be thus sounded alternately.

Fact (*b*) (*b*) If the same pair of sounds is concerned, whether pure tones or complex, any observer is consistent with himself at successive trials of (*a*), apart from irregularities of the kind that can be attributed to a finite sensitivity. If the pair are pure tones of the same frequency between the upper and lower thresholds of the observers, different observers agree amongst themselves to the same degree, though they may have different sensitivities. If the sounds are not pure tones of the same frequency, observers may differ regularly; older observers tend to find high notes less loud relatively to low notes than younger observers.

Fact (*c*) (*c*) Equally loud pure tones of the same frequency are equivalent in respect of all kinds of audition. Other equally loud sounds are not always equivalent. Thus of two equally loud sounds one may be more effective than the other in masking a third sound. Loudness, if it were defined with reference to simultaneous sounds, would not be the same thing as loudness defined (as we have defined it) with reference to successive sounds.

Fact (*d*) (*d*) Loudness is not additive. The operation most nearly satisfying the conditions for addition is simultaneous sounding. But in order that it should satisfy them, the following relations must be true:

If $X = X'$, then $X + Y = X' + Y$ ......(1).

If ϵ is such that

$\epsilon + X > X$ for any X, then $\epsilon + X > X$ for all X's ......(2).

Relation (1) holds if X, X' are pure tones of the same frequency, but not always otherwise. Relation (2) never holds; ϵ can always be chosen so that if it is sounded simultaneously with a comparatively faint sound the combination is louder than that sound, but that if it is sounded simultaneously with a loud sound, the combination is no louder than that sound.

The foregoing facts can be stated without reference to any magnitude or measurable property. The next two establish a relation between loudness on the one hand and, on the other, frequency and intensity, which are physical magnitudes measurable by processes universally accepted.

Fact (*e*)
k, p_k, x_k

(*e*) Any sound can be resolved (e.g. by Fourier analysis) into n components, of which any one k has the frequency p_k and the intensity x_k. The loudness of a sound is "determined by" these (p_k, x_k)'s. By that we mean (the definition must be noted) that two sounds having *all* their (p_k, x_k)'s the same are equally loud. On the other hand, sounds not having all their (p_k, x_k)'s the same may also be equally loud. The loudness of a sound generally, if not universally, increases with any x_k when the remaining x_k's are constant.

Fact (f)

(*f*) Among pure tones of the same frequency, loudness is determined (in the above sense) by their intensity only. The order of loudness is the order of the intensity. Fact (*f*) may be regarded as a special case of fact (*e*); but it is convenient to state it separately.

If we had an accepted "scale of loudness," we could assign a numeral X to represent the loudness of any sound. In virtue of (*e*) it would then be possible to find some relation of the form

$$X = a.F\,(p_k, x_k) \ldots \quad (k = 1 \text{ to } n) \qquad \ldots\ldots(3)$$

X, F

valid for all sounds and, in virtue of (*f*), a relation

$$X = a.F_p\,(x) \qquad \ldots\ldots(4)$$

F_p, x

valid for all pure tones of the same frequency p.

p

F must be single-valued, having the same value when all (p_k, x_k)'s are the same; F_p must be single-valued and monotonic. The problem of finding F and F_p would be in principle one of mere experimenting. It might turn out that F and F_p were not any known analytic functions; the relations might have to be expressed by graphs or mere tables.

All the scales of loudness that have been proposed in effect propose a function F or F_p such that the loudness of a sound can be measured (that is to say, a numeral X assigned to it) by means of a relation (3) or (4). One test that they must pass, if they are to be acceptable, is that when the measured arguments (p_k, x_k) for complex sounds are inserted in the proposed F, or the measured argument x for pure tones of a single frequency into the proposed F_p, then X resulting from (3) or (4) must actually be the same for all equally loud sounds.

If "equally loud" means "equally loud for all observers," no F can possibly pass the test; for when some sounds not of the same frequency are compared, all observers do not agree whether they are equally loud. The most that we can hope for is to find an F which will satisfy the test for some concordant group of observers. But for such a group the test is a real one, which some scales may satisfy and some not. The requirement that F shall have the same value for certain different sets of (p_k, x_k) imposes a very severe restriction on our choice of F. Thus it is agreed that the decibel-above-threshold scale does *not* pass the test for any important group; while the scale recently proposed by Fletcher and Munson* apparently does pass it within a certain range.

On the other hand not only can scales pass the test for F_p, but all proposed scales actually do so. Since there is only one variable, the only restriction imposed by the test is that F_p should be single-valued and monotonic. It would be satisfied if we simply put $F_p\,(x) = x$, and made our scale of loudness simply the scale of intensity. But everyone agrees that such a scale of loudness is not acceptable, even for tones of a single frequency. There must be some other test or tests which scales have to pass and our first business is to find out what they are.

---

* H. Fletcher and W. A. Munson, *J. acoust. Soc. Amer.* 5, 82 (1933).

One test may be that F_v must be the function to which an acceptable F reduces when there is only one component (p_k, x_k). One reason why the decibel scale is not acceptable, even for a single frequency, is doubtless that it does not pass this test. But this cannot be the only test. If it were, it would clearly be useless to try to establish a scale for a single frequency before we had a scale valid for different frequencies; yet there have been many attempts to do so. Again it seems clear that even our test for F is not sufficient. It is to be observed that if any F passes our test, so must any single-valued and monotonic function of F; yet all such functions of F cannot be equally acceptable. Whatever form of F we obtained in the first instance, it seems likely that *some* such function of it could be found such that, for a single component, it reduced simply to x. That function at least would not be acceptable.

This point is of great importance and must be illustrated by reference to the scales of physical magnitudes. The same magnitude can sometimes have several different scales, each of which is a single-valued function of another. Thus the transmission S and the photographic density δ of a plate, or the decay constant λ and the period T of a radioactive element, are the same magnitudes measured on different scales, related in the same way as any pair of proposed scales of loudness for a single frequency. Since $\delta = -\log_{10} S$ and $T = 1/\lambda$, the translation of a statement expressed in one scale into a statement expressed in the other is a mere algebraical transformation having no physical significance. It is a matter of mere convenience which we use. This is sometimes overlooked, because one form always is the more convenient; we are never tempted to use another.

On the other hand there are other magnitudes in respect of which it is not indifferent which of two scales related in this manner we use. The chief of these (possibly the only ones) are those which are additive and can be measured as A magnitudes. Thus it is not indifferent whether we write Ohm's law $E = a.I$ or $E^2 = b.I^2$; for the magnitude a (which is measured by this law as a B magnitude) is additive and measurable as an A magnitude also; but b is not. In this case there is a very definite reason for adopting one "scale" of resistance rather than another; one expresses, the other conceals, the supremely important fact of additiveness.

Now what is our reason for preferring one scale of loudness to another? Why are we quite sure that loudness ought not to be measured on a scale of intensity, but prepared to consider that it might be measured on the decibel scale by the logarithm of the intensity? Any statement in one scale can be translated by simple algebra into a statement in the other; and it does not seem that mere mathematical convenience is definite enough to make us so sure. There must here also be some facts revealed by one scale and concealed by another. It cannot be that which determines our choice in respect of resistances; for loudness is not additive. What is this fact? Is there indeed a real fact?

That is the main question discussed in this paper. We shall try to answer it by examining, under the headings set forth in the abstract, the scales that have been proposed. We shall ask what are the facts on which the scales are based and whether they really do provide an adequate reason for preferring one scale to another. In § 2 scales for pure tones of a single frequency will alone be considered; but the suffix to F_v will be dropped for brevity. Certain facts (*g*), (*h*), (*j*), (*k*), which apparently underlie these scales, will be assumed without the expression of any opinion whether they really are facts. Our question is only—If they are facts, do they determine a scale?

First perhaps we ought to say what we mean by a fact. We shrink from attempting any precise definition; but it will probably be admitted that the distinction between a fact and an arbitrary convention lies in its validity for all observers. In discussing sounds of different frequencies, it would perhaps be too rigid to insist that nothing valid for some observers and not for others is a fact; for, in virtue of "fact" (*b*), which is part of the definition of loudness, there could then be no facts and no scale determined by them. But in considering tones of a single frequency distant from a threshold, agreement in (*a*) and (*b*) is universal and it is permissible to be strict; since there are facts valid for all observers, we may demand that any other "facts" should be equally universal.

§ 2. TONES OF A SINGLE FREQUENCY

(i) *Mental estimates*

Fact (*g*)

Fact (*g*). If observers are told that a certain weak sound is to be represented by 1 and a certain strong sound by 100, they can be induced with a little persuasion to assign numerals by simple "guessing" to other sounds, especially those of intermediate loudness.

If their assignments were as instinctive and as concordant as those concerned in fact (*a*) they would deserve as much attention; they would lead at once to a relation (4). But then we should never have realized that there was any difficulty in measuring loudness. We might need a convention concerning the numerals to be assigned to the fixed weak and loud sounds; but if those could be fixed, no further question would arise; nobody would have any doubt what numeral was to be attached to any other sound. But the guesses are not as instinctive as judgments of equal loudness. Advocates of the method of mental estimates forget that the need for advocacy is the chief bar to its success.

Nor are they concordant. Superimposed on the irregularities of a single observer at successive trials are undoubted regular differences between different observers. Moreover it is at least highly probable that estimates are affected by the order of presentation of the sounds, and that a lower numeral would be assigned to a sound presented after a long series of sounds all louder than itself than to the same sound if presented after a long series all weaker than itself. From such discordant judgments a scale universally acceptable can be derived only by arbitrary conventions.

But though this method of obtaining a scale of loudness has never been considered seriously, it needs a little further examination, because it involves a question concerned also in our next method which is of much greater importance. How do we come to assign numerals at all by mere guessing and without actual measurement? We all constantly do it in every-day life; we talk of a 50-per-cent chance and of being twice as happy, when we have not carried through any process remotely resembling the measurement of any physical magnitude.

Surely the answer is that we are using some analogy between the circumstances of the judgment and those in which true measurement is possible. Thus at one extreme of this process of guessing is the "estimation of tenths." In such estimations we are undoubtedly using our familiarity with scales truly divided by actual measurement; if we had never seen such scales, the process would be impossible. Now, if we are using any analogy, the question arises whether the analogy is sound and whether the analogous facts are true. If there were some definite reason why the measurement of length broke down below the limit to which scales are divided, or if actually divided scales were entirely arbitrary and not based on true measurement, we should not attach much importance to estimation by eye. If that is so, before we accept mental estimates as true in the sense that the results of measurement are true, we should discover the analogy on which they are based and examine its validity very carefully. Universal agreement is no test of truth at all; for the most frequent source of universal error in the history of mankind has been an unquestioning acceptance of false analogies. And that is precisely what is most to be feared here.

(ii) *Equal relations*

It appears that, in addition to the judgments of equality of loudness concerned in fact (*a*), similar judgments can be made concerning equality of relations between pairs of sounds that are not equally loud. Expressed more accurately, the fact that we now add to our list is:

Fact (*h*) (*h*) Given any pair of sounds U, V of the same frequency and any other sound X, an observer can find a sound Y whose loudness is related to the loudness of X as the loudness of V is related to the loudness of U. X may, of course, be either U or V. All observers will agree in selecting the same Y within the limits fixed by their sensitivity. This cannot always be true when X is near the upper or lower limit of audibility; but it will suffice if it is true over a considerable range.

In order to discover how this fact may be used to limit further our choice of F, let us leave loudness for a moment and consider pitch, which is determined by frequency as loudness is by intensity*. Fact (*h*) *mutatis mutandis* is true of pitch. A and B define a musical interval; X and Y are tones separated by that interval. If we plot the frequency y of Y against the frequency x of X, we shall actually obtain a straight line,

a
$$y = a . x \qquad \text{......(5).}$$

If we took another pair of sounds U', V', defining another musical interval, and repeated the process, we should obtain another straight line and relation (5), but with a different value for a. Accordingly (5) sums up all the observations we could make in the matter in a single compendious formula.

---

* Pitch differs from loudness in that equality of pitch (which defines what we mean by pitch) can be determined by simultaneous sounding as well as by alternate sounding; but this is irrelevant.

Now these are precisely the circumstances in which, if we were making physical measurements (e.g. Ohm's law), we should say that we had found a new "derived" magnitude (e.g. resistance). It is therefore reasonable to say that we have found a new magnitude, pitch-interval, which is measured by a.

These facts impose a restriction on the forms of F in a relation $X=F(x)$, by which it is permissible to represent the relation between pitch X and frequency x. They show that our new magnitude a, pitch-interval, is determined (in the sense defined above) by the pitches of two sounds, the frequencies of which are related by the same straight line (5). That is to say, if we assume a relation $X=F(x)$, we must have

$$a=\phi\,(X,\,Y) \qquad \text{......(6)},$$

ϕ

where ϕ is a single-valued function and Y is the pitch of the sound whose frequency is related to the frequency of the sound of pitch X by (5). But this frequency is ax. Consequently we must have

$$a=\phi\,\{F(x),\,F(ax)\} \qquad \text{......(7)}.$$

(7) gives us a relation between the functions ϕ and F, but it still does not determine them uniquely. Thus we might have

$$F(z)\equiv Az^n;\quad \phi(u,\,v)\equiv(v/u)^{1/n} \qquad \text{......(8)},$$

where A and n are any real finite constants,

or

$$F(z)\equiv \log_n x;\quad \phi(u,\,v)\equiv n^{(v-u)} \qquad \text{......(9)},$$

where n is any real finite constant.

Now let us resume our discussion of loudness. We can again plot y against x (x, y being now intensities) for a given "interval of loudness" defined by A, B. But the graphs of y against x are not now straight lines; they are not any known algebraical curves. The graphs for different intervals of loudness, though similar, do not clearly differ from each other simply by the variation of a single parameter a. We cannot therefore measure intervals of loudness as we can measure intervals of pitch; and we cannot summarize all our observations in a single compendious formula, analogous to the equation $y=ax$ for pitch. We can represent them only by a family S of curves, each for a single interval.

We are therefore in an even worse position for choosing a scale of loudness than for choosing a scale of pitch; for the condition expressed by (7) is lacking. But we can try to regain condition (7) by assuming that the family of curves do represent a single function with a single varying parameter, although we cannot for the present determine it algebraically. This assumption (which will be called A) may be expressed by

Assumption A

$$y=S(a,\,x) \qquad \text{......(10)},$$

where S is an unknown single-valued and probably monotonic function. The

S

fact that the interval of loudness between two sounds is determined by a relation between their intensities is now expressed, not by (7) but by

$$a = \phi[F\{x\},\ F\{S(a, x)\}] \qquad \ldots\ldots(11).$$

Before we pass on, we must note that assumption A is precarious. We often discover empirically a set of relations that can be expressed by a family of curves, and sometimes later find the algebraic expression of those relations. It does not then always turn out that the curves differ in a *single* parameter; two or more parameters may be involved, which are not single-valued functions of each other*. If this should be the truth here, our subsequent procedure is unjustified. For if (10) should really be written $y = S(a, b, c, \ldots x)$, (11) no longer follows. However it will appear that the assumption will, in some measure, provide its own test.

But (11) does not in general determine F uniquely, any more than (7), even when S is known. In general, there will again be an infinite number of pairs of functions ϕ, F, satisfying (11). The form of S, known graphically, may exclude certain alternatives; thus, since the curves S do not now pass through the origin, solutions of the form (8) are (almost certainly) excluded from the start; but a new infinity of alternatives becomes open. The usual procedure at this stage is to assume a form for ϕ. If we make such an assumption we can now determine F graphically, using the graphically known S; the process need not be explained, because it has often been carried out and is familiar. If successful, it results in a graph $X = F(x)$ giving a relation between loudness and intensity, such that pairs of points on it whose ordinates are related in some manner determined by our choice of ϕ have the same interval of loudness and lie on a single curve of the original family. All the curves of that family are then summed up in this single graph; we have achieved the same unification as we achieved for pitch by means of the formula $y = ax$.

But it is to be noticed that the process might not be successful. It might turn out that if we used one of the original family of curves S, one graph F resulted, but that, if we used another, another graph F resulted. This possibility has generally been recognized; it has been realized that a scale of loudness determined by equal relations is legitimate only if all the curves S give the same graph F. If they do, the facts confirm assumption A, which is involved in the process.

If they do not, the procedure is unjustified unless there is one member of the family S so clearly distinguished from the rest that it is reasonable to select this member and neglect the rest. If this member is S_0, we can then replace (11) by

S_0

$$F\{S_0(x)\} = \phi[F\{x\}] \qquad \ldots\ldots(12);$$

again, if we choose ϕ we can determine F. Recent work† seems to suggest that there is such a unique relation, somewhat analogous to the octave in the field of pitch; just as untrained ears can identify an octave more definitely and certainly than (say) a tone, so they can identify one particular relation between loudness more definitely and certainly than any other. This fact will be called fact (j).

Fact (j)

---

[*] But it sometimes *does* turn out. Stefan's law and Planck's law were both discovered empirically as families of curves before their algebraic expression by functions with a single parameter was found.

[†] B. G. Churcher, A. J. King and H. Davies, *J. Instn. elect. Engrs.*, **75**, 401 (1934).

But whether we use fact (h) or fact (j), F cannot be determined without assuming the form of ϕ. Those[*] who have used fact (h) alone have assumed

$$\phi(u, v) \equiv v - u \qquad \text{......(13)},$$

which means that the loudnesses of equally related sounds are to be represented by numerals with a constant arithmetical difference. For example, the sound of loudness 100 must be related to that of loudness 90 as the sound of loudness 60 is to that of loudness 50. This assumption, which will be called B, is also implied in many discussions of the decibel scale; thus this scale has been criticized on the ground that the sounds represented on it by 100, 90, 60, 50 are *not* related in the manner just stated.

Assumption B

But it cannot be insisted too strongly that the assumption is not self-evident; it needs justification. The scale based on it will be determined by facts only if this justification consists in facts. Since nobody seems to have discussed the matter, we do not know what reasons would generally be alleged. But we can suggest three which might be alleged; it is to be noticed that, of the following reasons, (α) is based on facts, (β) is neutral with respect to facts, (γ) is contrary to facts. Here they are:

(α) It might turn out that, if any other form of ϕ were assumed, assumption A would not be confirmed; a different curve F would be obtained from different members of the original family S. Whether it does turn out so is a question of fact that does not seem to have been decided. Owing to the limited precision of the observations, it would clearly be impossible to prove that a slightly different form of ϕ would destroy assumption A; but if it could be proved that no form differing greatly from that defined by $\phi \equiv v - u$ is consistent with assumption A, there would be some reason for assumption B. But it would not be a very strong reason; for there is no *a priori* evidence for either A or B; it might be a mere chance that the errors in assumptions A and B compensated each other. If, on the other hand—as will probably prove the fact—other forms of ϕ are consistent with assumption A, there is evidence for A but none at all for B.

(β) Assumption B is more convenient in one way than any other. No other leads to a curve for F in which sounds standing in the same relation of loudness are represented by ordinates differing by equal lengths; it is much easier to pick out from a graph ordinates differing by equal lengths than ordinates related in any other manner, e.g. having lengths in the same ratio. But this, though highly convenient, is no more than convenient. Its convenience arises solely from the peculiarities of graphical representation, which cannot have any bearing on the facts of audition. So long as we are content to have an arbitrary element in our "scale of loudness" we are entitled to take advantage of this convenience; but we must not forget its presence and be surprised if other scales, based on other facts, do not agree with it.

(γ) What we have called an *interval* of loudness is sometimes called a *difference* of loudness; and pairs of sounds which lie on the same curve of the family S are said to show an equal difference of loudness. It is thence concluded that, if loudness is to be measured at all, such pairs of sounds must be represented by numerals showing the same arithmetical difference. This is a pure blunder[*], due to a confusion between the wide meaning of "difference" in ordinary discourse (when it means any relation of which non-identity is an essential element) and its narrow meaning in arithmetic (when it implies far more than mere non-identity). If "interval" is substituted for "difference," the fallacy appears at once; it becomes obvious that an interval may be represented by a constant ratio (as in the Pythagorean treatment of pitch) as well as by a constant arithmetical difference.

---

[*] We believe that this assumption has been made, but we cannot now trace the reference. All that is important to our argument is that it might be made; and this is clearly implied by the familiar criticism of the decibel scale, mentioned in the text.

Indeed there is a very definite reason for *not* identifying interval with arithmetical difference in both cases, and for *not* assuming the relation $\phi \equiv v - u$. The reason is that there is a very important class of magnitudes that are characterized by a relation for which arithmetical difference is the only appropriate numerical representation. These are the additive magnitudes, such as length, mass, and electrical resistance and capacity. The class and the relation are supremely important because, in virtue of the relation, magnitudes of this class and no others are measurable fundamentally —that is to say, without any other magnitude being measured first. If loudness were additive, none of the questions we are discussing would arise; there would be no question how it is to be measured; the only doubts remaining would concern "units" or (as we prefer to say) factors and standards. Assumption A may be permissible even if loudness is not additive, just as it is permissible for pitch, which is not even approximately additive. But it is unfortunate to select among the infinity of possible ϕ's that one which, if it were specially appropriate and distinguished from all the others, would make any selection unnecessary and the whole problem nugatory.

Assumption C

As we have noted, another possible assumption, which will be termed C, is that

$$\phi(u, v) \equiv v/u \qquad \text{......(14),}$$

so that the loudnesses of equally related sounds are to be represented by numerals in a constant ratio. Before we discuss it, we must insist that it is definitely inconsistent with assumption B; equal relations cannot be represented both by equal differences and by equal ratios; if we adopt C we must cease to expect equality in the relations of loudness 100 to loudness 90 and of loudness 60 to loudness 50. B and C can be true at the same time only if there are two distinct kinds of relations between loudness, one of which is appropriately associated with B and another with C. It is extremely improbable that there are two such kinds; and if they are, we must abandon for ever the hope of having a single scale to represent all the facts of loudness; we shall need at least two independent scales. Nobody who makes assumption C must criticize the decibel scale on the ground mentioned on p. 161.

Assumption C does not seem to have been adopted in quite the simple form in which we have stated it, but it is involved in the procedure of Ham and Parkinson[†] and of Churcher, Davies and King[‡]. They ask observers to pick out sounds which appear to them to stand in some prescribed ratio of loudness, such as $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{5}$. They then use (11) combined with (14), assigning the prescribed value to *a*. It is to be observed that this is equivalent to using (12) in the form $F\{S_0(x)\}=\rho.F\{x\}$, where ρ is the prescribed ratio.

---

[*] One of us has been guilty of it himself; see N. R. Campbell, *Proc. phys. Soc.* **45**, 590 (1933). The mistake is made easier by Fechner's errors, discussed on pp. 163–165.

[†] L. B. Ham and J. S. Parkinson, *J. acoust. Soc. Amer.* **3**, 511 (1932).

[‡] *Loc. cit.*

This procedure, though it might be well adapted to the establishment of a purely conventional scale, is ill adapted to discover facts. For the question which is put to the observer really consists of two parts: (i) Can you pick out pairs of sounds which appear to you to be equally related? (ii) If so, and if such pairs are to be represented by numerals in a constant ratio, what ratio do you prefer? If a definite ratio is named, the question substituted for (ii) is "leading," especially if fact (*j*) is true. For if there is a relation more definite than any other, an observer asked to pick out sounds "in a 2 : 1 ratio" will tend to select this relation, and will have to call it 2 : 1; no other alternative is offered him. The correct procedure to elucidate facts would be to separate the two parts (i) and (ii); it is then by no means certain that observers who all picked out the same pairs of sounds would all name exactly the same ratio; indeed experiments on mental estimates suggest strongly that they would not.

Even if they did, it would still be possible, as suggested on p. 158, that they were all using the same analogy and that this analogy might be false. Thus they might all pick out 2 : 1, because it is the "simplest" ratio between finite numerals and therefore appropriate to represent a simple relation in loudness. But the simplicity of 2 : 1 is based on the facts of counting, the determination of number, an additive magnitude having certain unique properties. The facts on which counting is based have no analogues in the field of loudness; an analogy based on simplicity is false. Again when they say that sound *B* is to sound *A* as 2 : 1, they may be judging that the relation between *B* and *A* is the same as that between *A* and some sound (e.g. silence) to which they instinctively attribute the numeral 0. But if that is so, they are simply blundering when they assign to *C*, which bears to *B* the relation of *B* to *A*, the numeral 4; they ought to assign the numeral 3. Indeed it is highly suspicious that the relation of 2 : 1, which plays so large a part in these investigations, is the most ambiguous that could be selected; no other leaves it so doubtful what numerals are to represent the higher powers, in the logical sense, of the relation.

We must not be misunderstood. We do not deny that by the procedure discussed here it is possible to obtain a scale of loudness which satisfies completely certain people and does not greatly offend anyone. What we do deny is that sufficient evidence has been advanced to prove that this concordance is derived from facts which must have universal validity, and not merely from generally accepted conventions and customs, the validity of which might be challenged by others with different training and habits. The importance of this distinction is discussed on p. 168.

(iii) *Just-perceptible differences*

We next turn to an entirely different procedure, now generally discredited, but historically important and still exerting a baneful influence. We shall assume the

Fact (k)

following fact (*k*). Given any sound X of intensity x, another sound Y can be found of greater intensity y, such that Y is louder than X, but any sound Y' of intensity between y and x has the same loudness as X. X and Y then stand in the just-perceptible-difference relation. We shall assume further that y is a single-valued and monotonic relation of x within a wide range of conditions, subject only to a determinate experimental error. This second assumption is by no means universally granted; some hold that y is affected profoundly by insignificant changes in experimental conditions, such as fatigue from previous observations. But, since the just-perceptible-difference method is clearly worthless unless the assumption is true, we shall make it for the sake of argument*.

δx

b

One way to use this fact would be to plot y against x and use the resulting curve in precisely the same way as one of the curves of family S in the equal-relation method†. The resulting method would be affected by the arbitrariness that was discussed on pp. 161, 163. But that is not the usual procedure. The usual procedure, due to Weber, is to plot $(y-x)$, which is usually called δx, against x or the mean of x and y. Weber suggested that the resulting curve was a straight line, so that

$$\delta x/x = b \qquad \text{......(15)},$$

where b is a constant. It is now agreed that his suggestion is false and that the curve is of some other form

W

$$\delta x = W(x) \qquad \text{......(16)},$$

where W is a function of which the algebraic form cannot be defined at present; but, if the facts are facts, (16) is a perfectly legitimate way of representing them.

δX

But the next step, taken by Fechner, is not legitimate. Fechner replaces b in (15) by δX, a constant difference in a magnitude X, loudness, which *ex hypothesi* he has not yet measured. He then writes $\delta X/\delta x = 1/x$, identifies $\delta X/\delta x$ with a derivative, applies to the equation the operator $\int_0^X dx$, and arrives at the equation

$$X = \log x - \log x_0 \qquad \text{......(17)},$$

giving a relation between X and x of the form $X = F(x)$ which we are seeking. His followers have applied the same procedure to (16), performing the integration graphically and arriving at

$$F \equiv \int_0^x \frac{1}{W(x)}\, dx \qquad \text{......(18)}.$$

It is as difficult to explain precisely why this is wrong as to explain why it is illegitimate to translate a French book with a German dictionary. The operator

---

* The question of fact, namely whether just-perceptible differences are unique, has become hopelessly entangled with the question of interpretation, namely whether Fechner's law expresses the fact. Those who deny the fact maintain, as a second line of defence, that Fechner's law does not express it; and those who see no fault in Fechner's argument are disinclined to admit that the facts he assumed are not facts. Here we want to separate the two questions completely.

† Indeed the curve might be the extreme member of that family; it would be so if all just-perceptible differences were recognized by the ear as similar intervals. It is doubtful whether they are so recognized. If the just-perceptible-difference curve were the extreme member of the family, it would provide a valuable test of the legitimacy of assumption A; but, as its membership may be doubted, the test is not stringent.

$\int_0^x dx$ is applicable only if the argument is a derivative, the limit of a ratio between two variables when the denominator tends to zero. But $\delta X/\delta x$ is simply not a derivative; δx is not a variable which can tend to the limit zero—it is always finite; and δX is, according to Fechner's assumption, not a variable at all but a constant. The whole procedure is mathematically absurd.

It is equally absurd physically; for even if an equation is mathematically integrable it does not follow that the equation resulting from its integration will be significant physically. An equation $y=f(x)$ states a numerical law only if the symbol = represents not only a numerical relation between the values of the magnitudes y, x, but also a physical relation between the systems to which they relate; generally they are magnitudes of the same system. If the equation is transformed mathematically it will acquire a new physical meaning only if this physical relation is replaced by some other. Whatever the physical relation, we can perform the mathematical operation $\int_0^x dx$ and arrive at a new equation

$$U=\theta(x) \qquad \text{......(19),}$$

where $\theta \equiv \int_0^x f(x)\,dx$; but it is not necessary that U should be a magnitude and (19) a numerical law. Thus if y is viscosity and x temperature, the integration will be physically meaningless; U will not be a magnitude and (19) not a law.

Indeed there are only two conditions in which integration is, in this sense, physically significant. The first arises when y is a derivative derived by differentiating (19); then integration will reverse the process and give the original equation. Thus y may be a coefficient of expansion with temperature and x temperature; (19) then gives the relation between the whole expansion and the change of temperature. But here y must have been, in effect, measured by using this relation. The second condition arises when additive quantities are concerned. Thus if y is the cross-sectional area of a tube and x the distance from one end, (19) gives the relation between x and U, the total volume from that end to x. But here $y.dx$ must be (or rather be proportional to) an additive magnitude, namely volume.

Neither of these conditions is present in the Fechner integration. dX/dx is certainly not a derivative obtained by differentiating some relation between X and x; for the argument starts from the hypothesis that none is known. And X is not an additive magnitude; for the existence of just-perceptible differences is inconsistent with additiveness; compare fact (*d*).

The preceding sentence is important, for an attempt has been made to avoid Fechner's mathematical absurdities by representing the just-perceptible-difference scale as one obtained by adding a number of equal steps from the threshold. But measurement by a number of equal steps is valid only if the magnitude is additive; a length can be measured by a number of equal steps, a density cannot.

The truth is that Fechner was simply wrong; his blunders have survived merely because text-book-writers copy one another*. The relation $X=F(x)$ obtained by his method is at best a very obscure way of representing the relation between δx and x, both of which are physical magnitudes and neither of which is conveniently identified with loudness.

---

* Similar criticism of Fechner's work has been made before. See, for instance, F. C. Bartlett, *Discussion on Audition*, p. 128 (Phys. Soc. 1931).

(iv) *The thresholds*

The best-known scale of loudness has been left until last because the facts that it expresses are much more limited than those on which the other methods depend. These other scales try to express relations of which any sound may be one term; the decibels-above-threshold method does not express any audible relation between two sounds; it expresses only a single fact about one single sound of each pitch, namely that there is a sound such that any sound of less intensity is inaudible. The assignment of numerals to represent other sounds depends wholly on physical relations, namely ratios of intensities, not on relations in which the ear is necessarily concerned (see p. 169). Such relations, involving all sounds, might have entered if it had been found that sounds of different pitch and the same distance-above-threshold were equally loud; but that turns out not to be true. Accordingly the decibel-above-threshold method expresses fewer facts than other methods and is to that extent more arbitrary. In its latest form, where the threshold is chosen arbitrarily and not as a result of experiment, it expresses no facts at all except those common to all these methods, namely that, among sounds of the same quality, loudness is a single-valued and monotonic function of intensity.

§ 3. SOUNDS OF DIFFERENT QUALITY

In this field the best that we can hope to achieve is a scale determined by facts valid for some "normal" group of observers; it would be impossible to obtain a scale applicable to frequencies as high as 10,000 equally valid for old men and children. But since this difficulty is overlain by others there will be no need to insist on it.

If we had concluded that some method determined uniquely a function F_p for a single tone, we should have had to inquire now whether, when p varied, the variation of F_p could be represented by the change of a single parameter; further, we should have had to inquire whether the loudness of complex sounds could be represented by some general function F of equation (3), which reduces to F_p when all components but one vanish. But since we have failed in our search for F_p we must proceed, if we can proceed at all, in the reverse direction. We must inquire whether the study of tones of different pitch and of complex sounds help us in our search for F_p.

It is possible that it might. Consider again the equal-relation method. It depends on assumption A, that there is a function $S(a, x)$ involving a single parameter for sounds of the same pitch. If that assumption were fully established for all pitches, it would be natural to suppose that this parameter might be a function of pitch p, so that all the families S for all pitches could be represented by a single formula

P

$$y = S\{P(p), x\} \qquad \ldots\ldots(20),$$

where P must be single-valued but need not be monotonic.

Assumption D

It would not be as easy to produce indirect arguments for this assumption, which may be called D, as for assumption A; it seems to require that an algebraic form for $S(a, x)$ should be found. But if it could be established, interesting consequences might follow. From (11), on substitution of $P(p)$ for a, pairs of functions ϕ and F might be formed as before; but since the conditions to be fulfilled (if assumption D had already been established) would be more stringent, the number of permissible pairs would be reduced. It *might* turn out that only one pair was permissible; then $X=F(x)$ would be a true measure of loudness, free from arbitrariness. Whether it would turn out so is a question of fact, which it is impossible to investigate until assumption D has been established; but this appears to be the direction in which it seems most profitable to seek a further advance by means of the equal-relation method.

What may seem at first sight to be an alternative method is that of Fletcher and Munson*. The problem with which they are concerned primarily is this. Given all the (p_k, x_k)'s for a sound, find by calculation the intensity x_r of a reference sound of fixed frequency p_r which is equally loud. Any solution of that problem—and we assume that Fletcher and Munson have solved it—requires the finding of an F which satisfies the test given on p. 155 and has the same value for different sets of (p_k, x_k) which correspond to equally loud sounds of different quality. But the solution does not require that the F found should satisfy the tests given on p. 156.

As a matter of fact, Fletcher and Munson's solution does not satisfy the test that F reduces to F_p for a pure tone. If this were all, perhaps the conclusion should be that this test is too strict and that "reduction" in the mathematical sense is not necessary; all that is required is that a set of functions F should be set up covering between them the whole field of pure and complex tones. We are not entirely ready to admit that conclusion; it seems to us that, if the scale is to be satisfactory, the loudness of a pure tone of intensity x ought to be calculable from the loudness of a pure tone of the same frequency and intensity 1; or, in other words, that the tone of intensity x ought to be capable of being regarded as a complex of x pure tones of intensity 1. For if that is not so, the scale does not express the very important fact that intensity (unlike loudness) is additive. It seems impossible that this condition should be fulfilled unless F reduces mathematically to F_p.

But that is not the only reason why Fletcher and Munson's scale does not solve the problem with which *we* are concerned. Their F_p is simply decibels above a given arbitrary threshold. If they had shown that it was impossible to arrive at any function F, passing the test given on p. 155, unless F_p had this form, they would have produced a fact affording a reason for adopting their scale F_p and therefore F. But it does not seem that they have shown this. So far as we can ascertain, they could have carried through precisely the same procedure if they had made F_p for the reference tone simply the intensity x_r. Of course they would have arrived at a different F; but they would have arrived at some F valid over the same range. If that is so, they have produced no reason such as we are seeking for preferring the scale F_p, which they adopt, to a scale in which, for pure tones, loudness is simply identified with intensity.

---

* *Loc. cit.*

There is then no evidence that we can find in the study of complex sounds facts which will serve to select a single scale of loudness from the infinite number, all single-valued functions of each other, consistent with facts (*a*)-(*f*). We still have to admit that all scales contain an arbitrary element at present irremovable.

§4. GENERAL CONCLUSIONS

Are we then to conclude that the scales are all useless? This question must be relative to some purpose. One of the chief aims of pure physics is to arrive at theories to explain facts. In the course of its long history, it has always turned out that theories can be found only when the facts, if expressed numerically at all, either are expressed in terms of "true" scales, based completely on facts except for an arbitrary factor, or else are equally expressible on all scales alternative according to the principle laid down on p. 156. Why this should be so is a mystery, but there are abundant examples; for instance, there could be no adequate theory of heat before the absolute scale of temperature was developed.

If then our purpose were to develop a fundamental theory of audition, we might well despair of making an advance so long as arbitrary scales are necessary and no others are available. Or again, if our purpose were merely to describe the facts in a compendious form, and if the purpose could not be attained at all unless all the facts were described in a single compendious form, we might equally despair; for in physics the only laws that describe all the relevant facts are those which are found to have a theoretical basis. Empirical laws are always of limited scope.

But that is not the position. A scale of loudness is required by the engineer concerned to reduce the annoyance caused by his machinery; by the designer of radio-receiving sets concerned to obtain faithful reproduction of music; by the telephone engineer concerned with masking of speech by interference. The first is concerned mainly with the upper range of audibility and with complex sounds, the second with the lower range and with nearly pure tones, and the third with the lower range and complex sounds. Their fields are so distinct that they very seldom need to correlate their observations. There is no reason why they should employ the same scale. While each believes that the scale that suits him is a "true" scale, imposed by ineluctable facts, they will never agree to differ; each will try to convert the other to what he believes to be the one and only truth. But if they would realize that all scales are arbitrary and that one is just as little true as the other, toleration would become easier. Whether any of the scales that have been proposed is actually well adapted to any of the various purposes for which scales are required, and, if so, which scale is adapted to which purpose—these are questions on which we express no opinion. But we cannot help believing that if the latitude which the presence of an arbitrary element permits were generally recognized, progress towards the adoption of standard scales, each for its own particular purpose, would be more rapid.

There is a final question that we are loth to attempt, but is almost sure to be raised in any discussion of this matter. Is there any reason why it should be impossible to find a true scale of loudness?

The only way to answer this is to examine the differences between physical magnitudes which can be measured and psychological magnitudes which, we feel, ought to be measurable, but which actually are not. The difference that appears to us most distinctive is this. Physical magnitudes are measured by processes depending on sensory judgments which can be made with any sense organ. Sight is usually employed, but that is merely for convenience. Given suitable apparatus a blind man could determine for himself a weight, a density, or even magnitudes closely connected with vision, such as wave-length or intensity of radiation. But it is inconceivable that he could measure colour or brightness. These terms depend for their meaning on visual judgments; anything that a blind man could measure without consultation with a sighted man would, *ipso facto*, not be colour or brightness. Psychological magnitudes are those which, if they are to be measured at all, must involve judgments made with a single sense organ. Loudness is one of them, for its meaning is derived from fact (*a*).

An equivalent but cruder way of expressing the difference is to say that the physical magnitudes are properties of all matter, while the psychological magnitudes are properties of the particular pieces of matter that form our sense organs. If that is so, some explanation appears of the fact that psychological magnitudes do not possess the properties that would make them measurable. They are not additive because our sense organs are not constituted in such a way as to make them so; additive mechanisms are generally very simple and our sense organs are very complicated. Again they do not appear, generally, as constants in a numerical law between physical magnitudes, because the only laws in which they could appear as constants would be those involving the operation of our sense organs, and the physical magnitudes characteristic of these organs are not open to observation in the living and conscious subject.

Further we can see why some psychological magnitudes such as musical interval, discussed on pp. 158, 159, are measurable as *B* magnitudes. The reason is that they do not involve this structure so intimately as the others. They are all of the nature of differences; the difference between the effects of two stimuli on a mechanism may be more nearly independent of the structure of that mechanism than the effects themselves; it may be more nearly a property of matter.

Does this mean that, if we knew more about the operation of our auditory organs, we could hope to measure loudness on a unique scale? Our answer is negative, so long as the general outlook on the problem remains what it is at present. Loudness can never be measured as an *A* magnitude, because it is not additive. That is a fact which nothing can change; it would not be changed if we found that loudness was intimately connected with some *A* magnitude characteristic of the auditory organs, such as the rate at which impulses pass along some nerve system. The most that any further knowledge of these organs could do would be to correlate loudness with some magnitude *H* characteristic of these organs, so that the order of loudness would be the order of *H*. If *H* were a constant in a numerical law of these organs, loudness could then be measured in the scale of H as a B magnitude. But it does not follow that this scale would be satisfactory; it might be unsatisfactory just as the intensity scale is unsatisfactory for pure tones, although for them intensity is correlated exactly with loudness. If it were already satisfactory, we should have made no advance; if it were unsatisfactory, it would not become satisfactory because we had discovered that H and loudness were correlated.

This is the crux. At present the reasons that determine whether a scale is satisfactory are based on facts (*g*), (*h*), (*j*). If these do not determine the matter, doubt can be removed either by finding other facts of the same kind, or by ceasing to attribute importance to these facts and being as indifferent to scales as we are in respect to some physical magnitudes. There is no third choice. It is perhaps too early to despair of the first; but in case it proves ultimately impracticable, it is worth while to point out that the second is open. If we could only rid ourselves of the feeling that loudness ought to be additive and that equally related sounds ought to be represented by numerals in some simple arithmetical relation, disputes concerning scales of loudness would vanish; the only remaining problem would be to measure the loudness of complex sounds on some scale, no matter what. We should concentrate on F and cease to be interested in F_p, which might be any single-valued function of intensity. Then it would be quite possible that the best way to find F would be by means of H as intermediary. Here we find ourselves in complete agreement with what we understand to be the view of Fletcher and Munson, even if we cannot accept all the arguments by which they support it.

§ 5. ACKNOWLEDGMENT

We desire to acknowledge the help we have received from discussions with our colleague, Mr D. A. Oliver.

DISCUSSION

Dr J. H. SHAXBY. Two papers in this month's *Proceedings of the Royal Society* B describe experimental investigations made recently by F. H. Gage, in this institute*, on the possibility of forming scales of sensation. The sensations chosen were brightness and loudness, and the conclusion reached in both cases is that the method of dividing a given sensation-interval into two equal parts does not lead to a consistent scale. Briefly, if we thus bisect the whole interval and then each of its halves and finally similarly divide the interval between the ¼-way and ¾-way points, we do not arrive again at the same stimulus as that for the original ½-way point. The term bisection is not to be taken as implying equality of differences of loudness or brightness, but only equality of the two "intervals" in the sense in which Campbell and Marris use that term. This implies the failure of other methods of founding scales, such as the just-perceptible-difference step method, since the number of steps to bridge the interval to a brightness (for instance) lying half-way between two given brightnesses differs according to the means adopted for defining the stimulus giving the bisecting brightness, and this is indeterminate.

---

* The Physiology Institute, University College, Cardiff.

A fact that emerged in the course of the work is of great interest. Many observers have in different researches halved sensations, as for instance in Churcher's paper referred to by the authors and in older work on brightnesses. We tried a somewhat different procedure, namely the bisection of the interval between a moderate brightness and a very feeble one, repeating the tests with a progressively weaker dim light until the threshold was nearly reached. The limit of this, the bisection between the fixed brightness and zero brightness, would, one might have supposed, have been equivalent to halving the fixed brightness. On the contrary, the bisecting stimulus was but little above the feeble stimulus, and seemed to approach it as the latter approached the threshold. It is clear that the judgment made in this way is quite different from that in the halving experiments in which the influence of the comparison weak light is lacking.

The observation is further evidence against the relation $\phi = v - u$, and possibly, though not certainly, in favour of $\phi = v/u$.

While all who tried Gage's bisections agreed that the judgment was definitely possible and comparatively easy to make, none felt any certainty as to what they were doing. Whether they were obtaining equal differences or equal ratios of loudness or brightness they did not know: all they were sure of was that there was some definite equality involved. In the authors' phrase, the two intervals were equal. My own impression was that it was *not* the differences which were equal, but that the difference between the greatest and medium sensations was larger than that between the medium and smallest, although the intervals were equal.

The authors are not the first to demolish the mathematics of the Weber-Fechner law, but they have done it with unusual thoroughness. They do not add the comment, itself not new, that an equation expressing the relation of sensation to stimulus as a continuous function and derived from another equation formally asserting the discontinuity of that relation, is indeed remarkable.

But despite the mathematical illegitimacy of its origin the law does express, what the decibel scale of course also expresses, the fact that sensations seem to run more or less parallel to the logarithms of the energies of the stimuli which evoke them, while commonsense immediately rejects the suggestion that their values are proportional to the energies themselves. This surely is the fact, revealed by one scale and concealed by the other, which has given such undeserved prominence to the decibel scale and misled so many into regarding it as a scale of sensation and not merely as one of stimulus.

Dr J. E. R. CONSTABLE. One small point may perhaps be added to the valuable analysis which the authors have given. I refer to their suggestion that the loudness-doubling upon which other authors' work is based is probably only a "most-easily-recognized interval". It has been argued that the agreement between loudness-doubling results and the monaural-binaural measurement is an argument in support of the possibility of recognizing a factor of 2 in loudness. The experiments undoubtedly show that a sound when heard by two ears is perceived as louder than when heard by one. It does not appear, however, to have been demonstrated that the loudness is, in fact, twice as great in the former as in the latter case. Might it not be possible that the agreement between the experiments is due to the fact that the difference between monaural and binaural listening is a relation with which everyone must be familiar and might merely constitute the most-easily-recognized interval to which the authors refer?

As regards the practical aspect of the authors' results it may be remarked that even if the supposed loudness-doubling is only a most-easily-recognized interval, the usefulness of loudness scales such as that put forward by Churcher, King and Davies would not necessarily be impaired; for, from the practical point of view, a scale made up of a series of most-easily-recognized intervals is as significant as one based upon a succession of loudness-doublings.

Mr C. R. Cosens. An alternative to the procedure outlined in the second paragraph on page 163 suggests itself, which would appear to be free from the leading-question objection. Let there be two sounds, A and B, whose loudnesses are a and b respectively. The observer is required to determine what, in his opinion, is the value of the ratio a/b. Provide two adjustable sources of sound X and Y (say loud-speakers fed from potentiometers) with an arbitrary calibration of their volume-controls (say the angle θ through which the control-knob is turned).

The observer first adjusts the knob of X until it appears to give the same loudness as A, and $\theta (a)$ is recorded. He is then asked to adjust Y until it appears twice as loud as X, and $\theta (2a)$ is recorded. X is then adjusted to be twice as loud as Y, giving $\theta (4a)$. This doubling process is repeated until the sound is undoubtedly louder than B, and a graph can then be plotted connecting loudness as estimated with the arbitrary variable θ.

Sound B is then introduced, for the first time, to the observer, who is asked to adjust one of the sources, X or Y, until it is of the same loudness as B, and $\theta (b)$ is recorded. The value of the loudness of B in terms of multiples of a can then be interpolated from the graph.

A few rough experiments were made on these lines some years ago, the tuning-note of the B.B.C. and a wireless receiver being used.

θ was the reading of a square-law voltmeter connected across the loud-speaker windings, the volume was adjusted to a definite θ in each case, and observers were asked to turn up the volume-control until the sound appeared to them to be twice as loud. Results of course varied considerably, but a more or less linear relation with θ was shown, leading to the conclusion that "twice as loud" meant four times the voltage across the loud-speaker, i.e. 16 times the power expended, so that for this particular pure note and starting from one particular loudness, the mean observer considered the loudness to vary as the fourth root of the power converted into sound.

Dr R. T. Beatty. During the year which has elapsed since Dr Campbell read a paper on the measurement of sensation before this Society he has made a psychological advance, for the operations which he then described as "guessing" he now promotes to the rank of "mental estimates," a term which is less suggestive of a game of pure chance.

The available experimental evidence as to subjective loudness may be summarized in a single sentence as follows. *In the domain of sensations of loudness the brain can perform the operations of multiplication and division.*

Four investigations* are to hand in which observers were asked to alter the loudness of a sound to some multiple or sub-multiple of this loudness. For this operation they depended solely on their own subjective intuitive judgment as to what the magnitude of the change in loudness might be. When the four curves are compared the maximum deviation from the mean curve does not exceed 8 decibels in a range of 80 decibels, i.e. an intensity-ratio of 6 : 1 in an intensity-range of a hundred million to one. If this is guessing, the guesses are remarkably consistent. They are further supported by a deduction from the work of Fletcher and Munson†, a deduction which Dr Campbell has failed to make, but which has now been pointed out by Mr Churcher. A sound at a level of 100 decibels applied to one ear seems of the same loudness as a 90-decibel sound applied to both ears. On the assumption (derived from consideration of the numbers of nerve impulses) that binaural hearing gives twice the loudness of monaural hearing, we thus find that a reduction of 10 decibels gives a loudness-ratio of 2 : 1, and so a (stimulus, sensation) curve can be plotted. The agreement of this curve with the curve deduced from mental estimates is remarkable. The ratio 2 : 1 has now hurled itself so violently into the evidence that it is difficult to ignore it.

But the authors simply cannot understand how the idea of number can enter into sensations of loudness. I had hoped that my remarks‡ on the 1933 paper might furnish a clue. It is unnecessary to repeat the suggestion then made except to say that it indicated that the correlation between eye and hand involved in handling weights, together with the linear relation between weight and afferent nerve impulses, furnish a numerical relationship which might be transferred to the auditory or visual receptor system.

Several attempts have been made to prove that Fechner's law is mathematically unsound, but if our sensations were additive Fechner's law would be valid, and indeed the analysis given in this paper can by a slight modification in the wording be used to prove that it *is* valid. This paradox is due to the fact that Fechner's integration is not a pure mathematical construction. Its history serves to warn us that the algebra of sensation is not necessarily the same as that of physical magnitudes.

It is difficult to discover what, apart from conclusions of despair, the authors' general conclusions are. One remark however, to which they seem to attach some importance, deserves comment. They say, "If we could only rid ourselves of the feeling...that equally related sounds ought to be represented by numerals in some simple arithmetical relation, disputes concerning scales of loudness would vanish."

* Richardson and Ross, *J. gen. Psychol.* p. 288 (1930). Ham and Parkinson, *J. acoust. Soc. Amer.* **3**, 511 (1932). Geiger and Firestone, *J. acoust. Soc. Amer.* **5**, 25 (1933). Churcher, King and Davies, *J. Instn elect. Engrs* (October, 1934).

† Fletcher and Munson, *J. acoust. Soc. Amer.* **5**, 82 (1933).

‡ Beatty, *Proc. phys. Soc.* **45**, 572 (1933).

Why should we try to rid ourselves of this feeling, since five investigations* are on record to show that such numerical relationships actually exist?

Mr W. F. Floyd. I cannot agree with the desire of the authors, expressed by one of them (G.C.M.) in his introductory remarks, to place this discussion outside the field of psychology. Loudness as defined in the paper in (*a*), or as defined in any admissible way, is essentially a mental estimate. The estimate may be a psychological correlate of the physiological change resulting from the stimulation of the outer ear, or it may be a simply definable function of the endo-somatic response dependent on whether we reject or accept a purely behaviourist (physico-chemical) description of human effort. Mental estimates are acceptable, statistically, to psychologists and form reliable measures, provided that the usual statistical conditions are satisfied, namely that the sample is large and random and the experimental methods do not arbitrarily bias the results. If the authors believe that human effort may be explained in terms of physics and chemistry only, their problem is proper to the physiologist and may be solved by experiments in which a measure is made of the potential-fields and barriers in the nervous system and the cortex†. If they believe in mind they must accept mental estimates and the statistical methods of the psychologist, and use the mean value of mass data, with the variance, to set up a loudness scale and so determine their function (3).

The difficulties which the authors conceive in the construction of a scale for pure tones of a single frequency by the method they designate "mental estimates" are difficulties of experimental technique only, which occur in all psychological experiments. It appears to me, as I interpret the authors' statements, that the method § 2, (i) which they would discard is based directly upon their first fact of audition, (*a*). Indeed, when the experimental technique is correct, the method is a restatement of the last sentence of (*a*).

In place of equation (16) of the paper I would suggest the relation

$$\Delta X = \phi\,(x, \Delta x) \qquad \ldots\ldots(16a).$$

This does represent the facts, subject to the condition that ϕ is monotonic and single-valued. To every variate of ΔX there correspond unique variates of the variables x and Δx, and we have thereby avoided the error of assuming that all the ΔX's are equal, an error which the authors condemn but which they commit on p. 165 when they say "The preceding sentence is important, for an attempt has been made to avoid Fechner's mathematical absurdities by representing the just-perceptible-difference scale as one obtained by adding a number of *equal* steps from the threshold."

The term "equal" is applicable where two stimuli, occurring alternately, arouse physiological changes which have undistinguishable psychological correlates. The term must be warily applied to increments in perceptions. A few experiments in this direction have shown the exponential relation between stimulus and percept to be approximately true for the subjects examined.

---

* See footnotes *, †, on p. 173.

† *Vide* the work of Adrian, Lashley, and others.

Mr B. G. CHURCHER. In the introductory part of the paper the authors have performed a useful service in laying down with precision certain fundamental propositions, with which most acoustical technicians will be, I imagine, in complete agreement. The realization that loudness is not additive eliminates the necessity for considering two scales, viz. the decibel loudness scale and the scale based on the sum of the observed just-perceptible increments of intensity. As the authors point out, it is theoretically possible to use many types of scale. But, from both scientific and utilitarian points of view, it is desirable to use a scale in which the numerical value assigned to a sound is in direct proportion to the loudness-sensation evoked by that sound. If such a scale is possible, then to use any other kind is to create unnecessary complication. To obtain such a linear scale has been the object of my colleagues and myself. Since most noise-measurements are stated in terms of the intensity of the equally loud reference tone above threshold intensity, the relation between decibels above threshold and loudness for a pure 1000-cycle tone is of particular interest. Since loudness is a sensation, the fundamental criterion of relative loudness is a representative individual's assessment of it. Hence the method of mental estimates of relative loudness is an important guide even if it is not the most precise method of arriving at the (decibel, loudness) relation. Loudness scales which are in conflict with the data of representative mental estimates are open to suspicion. The authors have apparently misunderstood the method of mental estimates employed by my colleagues and myself and I do not think that their objections hold. If the five investigations that have been carried out on this subject* be compared, it will be found that, one being omitted, a representative curve may be drawn relating decibels and loudness on a scale, such that the loudness evoked by 100 db. above threshold is designated 100 and that any individual investigation deviates from the curve by not more than 8 db. (A lantern slide was shown.) An explanation of the faculty of making intuitive estimates of relative loudness is afforded by the nerve-impulse theory of audition†, which also explains the phenomena observed when a sound of given intensity is applied to one ear and alternatively to both ears. The theory indicates that the loudness is doubled in the latter case. If the relative intensities for monaural and binaural listening for equal loudness are measured, a (decibel, loudness) relation can be deduced. Data from two independent sources show good agreement and a representative curve can be drawn. (A slide was shown.) Fletcher and Munson‡ have proposed another method based on their theory of the loudness of complex sounds. But the general validity of their theory is doubtful as no distinction has been made between sounds with harmonically related components and those with non-harmonically related components. The enhanced loudness obtained with harmonic relations is beyond doubt and has been recently confirmed by Fletcher himself. Tone-combination methods are not, therefore, admissible for the construction of a loudness scale. If now the representative mental estimates and monaural-binaural curves be compared, close agreement over a loudness-range of 200 down to 0·4 will be found. (A slide was shown.) The final curve based on these results is quite well represented by the law $L = d^5 \times 10^{-8}$, where L is the loudness and d the decibels above threshold. (A slide was shown.) The concordance of the nerve-impulse hypothesis, the mental estimates and the monaural-binaural results seems to indicate that a representative, if not a unique, loudness scale can be formulated.

---

* L. F. Richardson and J. S. Ross, *J. gen. Psychol.* **3**, 288 (1930). D. A. Laird, E. Taylor and H. H. Wille, *J. acoust. Soc. Amer.* **3**, 393 (1932). L. B. Ham and J. S. Parkinson, *J. acoust. Soc. Amer.* **3**, 511 (1932). P. H. Geiger and F. A. Firestone, *J. acoust. Soc. Amer.* **5**, 25 (1933). B. G. Churcher, A. J. King and H. Davies, "The measurement of noise with special reference to engineering noise problems," *J. Instn elect. Engrs* (October, 1934).

† Dr R. T. Beatty's remarks: "The measurement of visual sensation," Discussion, *Proc. phys. Soc.* **45**, No. 249, p. 565 (July, 1933).

‡ Harvey Fletcher and W. A. Munson, *J. acoust. Soc. Amer.* **5**, 82 (1933).

Mr W. West. Experience with subjective tests (i.e. tests of what the authors term "psychological magnitudes") has led me to regard these as comparative only, requiring skill to ensure that the comparison is both apt and fair. I am therefore surprised to learn (p. 168) that telephone engineers and others believe loudness scales to be "true" scales. The most universally employed loudness scale is, as I regard it, no more than a means of ascribing numerical values to one aspect of noisiness. It has been given the name "equivalent loudness" in order to avoid confusion with the general use of the word "loudness." It equates, by subjective test (comparison), the loudness of a noise with that of a standard tone (arbitrarily selected as 1000 c./sec.), and states the equivalent loudness of the noise in terms of the sound pressure of the standard tone (an exact physical measurement). It seems to me that this method is referred to, but misrepresented, in paragraph 2, (iv) on p. 166. Details of standardization of this scheme may not yet be complete, but the principle is very generally accepted by engineers who have to measure noises. I would refer the authors to some tests and theories of loudness by U. Steudel*.

Mr A. J. Aldridge. The measurement of loudness is becoming increasingly important and obviously some scale is necessary. The authors have investigated various suggestions, and, so far as I have followed them, have indicated that no present method is satisfactory. They appear to doubt whether any true scale is possible. If this be so, the only thing which engineers can do is to carry on with the more or less arbitrary scales that have been selected as most suitable for the particular purposes they have in hand. On the other hand the authors, after having made a detailed examination of existing scales, and incidentally having pulled them to pieces, should be in a good position to suggest some improvement. I think we are entitled to ask them to do so.

Mr J. Guild. The authors have given a very clear treatment of the problem of loudness-measurement, as that problem appears to be envisaged by the majority of acoustical engineers. There are some points in the development of their arguments with which I do not completely agree, but as none of these leads to any material difference in the significant conclusions I will abstain from discussing them.

A somewhat startling conclusion which has emerged from the discussion which has followed the paper is that workers on acoustical problems, in so far as they were

* *Hochfrequenztech. u. Elektroakust.* **41**, 116 (1933).

represented by the participants in the discussion, hold views which are essentially different from those of Dr Campbell on the kind of association between number and magnitude which constitutes measurement. If the theory of measurement depended in any way on a specialized knowledge of each field of phenomena in which it is applied, one could hardly escape the conclusion that with this weight of specialist authority against him Dr Campbell would be well advised to revise his ideas; but the principles which govern the association of numbers with magnitudes in order that the former may be *measures* of the latter and not merely more or less arbitrary numerical tickets, do not depend on the nature or properties of the thing measured. They depend simply on the nature of measurement, and it is difficult to believe that any physicist does not accept as fundamental the general principles on which the arguments in the present paper are based.

Mr L. V. K. REIN. It seems advisable to record an aspect of the work that has reached, to all intents and purposes, international acceptance. The particular field is the measurement of hearing-loss on the sensation-unit, decibel scale, as laid down for use in connection with the 2-A and 2-B (Bell Laboratory) audiometers. Over a steady series of researches carried out in this country and America during the last twelve years, by otological and research groups, it has been definitely established that not only may an observer's hearing-capacity be charted, but such progress has been made with the principle that libraries have been laid down where tens of thousands of hearing-test charts are now kept, and characteristic audiograms for use in diagnosis have been conclusively arrived at as a result of these investigations.

The concrete application of the principle has been further established as a result of the recent two years' Hearing Aid Research Survey, carried out under the auspices of the Ministry of Health, whereby a series of instrument-specifications based on performance curves calibrated in a state laboratory made possible the practical application of the work. A series of these instrument-performance curves were converted to the sensation-unit scale, so that it became possible to match the indicated hearing-loss curve with an instrument-performance curve compensating the patient's defect with an accuracy proved to be useful.*

As a result of careful comparisons, results have proved that a working co-operation with otologists using the half-intensity-unit (speed of decay of a tuning fork) method of testing is simple and satisfactory, and as a result of steady investigation where hundreds of cases have been tested by both fork and audiometer methods, provided that tests are carried out under silence-room or equivalent conditions, the discrepancies have been found to be practically negligible.

In the light of the vast field covered by the subject, an abstract taken from the conclusions arrived at during a similar discussion held recently by the German group in Berlin seems to express the present trend to reach a general agreement of international understanding in setting up an agreed unit and measurement, and whilst it is the writer's endeavour to take an unbiased view, the results of our practical principle are proving of such value that there is little likelihood of an immediate need for the re-establishment of a further unit in connection with our branch of the work, until an extensive series of investigations and experiments have been completed along the lines of those already in progress. These investigations seek to establish what has up to the present been a much neglected subject, the percentages of comparisons between an individual's hearing by air conduction in relation to his bone-conduction hearing-capacity.

---

* H. M. Wharry, *The Practitioner*, **129**, No. 5.

A series of routine tests are being carried out at the present time in conjunction with three of our hospital research groups, and it is to be hoped that a new insight into an observer's ability to detect directional sounds may be established, even in the instance of a person with normal hearing, when a series of group categories based on the combined bone-conduction and air-conduction hearing curves have been made.

The object of bringing these facts to light is to eliminate any prospect of an erroneous idea being formed in connection with the measurements of hearing-capacity by established sound-intensities over a series of frequencies, and to offer the assurance that data are to-day available which prove beyond any question of doubt that the unit adopted in connection with our own field of work is adequately suitable, and offers a method of dealing in an entirely satisfactory manner with an individual's hearing-capacity, by the use of a measured scale which is now accepted and in use in centres and hospitals in the majority of cities throughout the world. The data, performance curves, and some thousands of audiograms taken in conjunction with otological and research groups, together with records of the work accomplished in America and Germany in this field, have been compiled and it is to be hoped that, when a little further investigation dealing with the aspect of bone-conduction hearing has been reached, full publication will be possible.

Mr H. DAVIES. In considering the method of mental estimates, § 2 (i), the authors have raised the question of the nature of the criterion involved in the decision that one sound is, say, twice as loud as another. It has been suggested by Dr R. T. Beatty* that for a pure tone the criterion is the number of afferent nerve impulses per second evoked by the stimulus and that we assign numerals to auditory sensations by analogy with the nerve impulse frequencies involved in the more familiar experience of lifting weights.

The idea is very suggestive and in a paper which is to be published within a few days† I have shown that if this hypothesis is accepted it is possible to construct a scale of loudness from the data of monaural and binaural minimum-perceptible changes of intensity by a method quite different from any of the four considered by the authors. Although it uses the minimum-perceptible-change data the method does not, of course, involve the process of integration of minimal changes to which the authors have administered so effective and so well merited a castigation.

In § 2 (ii) the authors have instanced a recent publication of Messrs Churcher and King and myself as indicating that there is a unique loudness relation, analogous to the octave, which is more readily identified than any other by the untrained ear. Of this I am not convinced. The work referred to is not entirely conclusive on this point, since the only ratios employed were 2 : 1 and 4 : 1, and if there is any particular facility associated with the octave relation it will presumably also be involved in some degree in the double octave. Fact (*j*) therefore still remains open to question, but on the ground of introspective experience I doubt whether it is true, except in so far as 2 : 1 is the simplest and most easily appreciated numerical ratio. The authors discuss the use of the 2 : 1 ratio and they suggest that its simplicity is relevant to counting but that "the facts on which counting is based have no analogues in the field of loudness." But if, as has been suggested, the estimation of loudness is essentially an estimate of nerve-impulse frequencies, then the process *is* analogous to counting and the simplicity of the ratio 2 : 1 has significance.

---

* *Proc. phys. Soc.* **45**, 572 (1933).

† *Phil. Mag.* (Nov. 1934).

It will be agreed that in making an estimated partition of a straight line ruled on paper the easiest division is into two equal parts and that the difficulty of the operation increases with the complexity of the divisional fraction. Considering only the mental state of the subject making loudness-estimates, I rather think that the facility conferred by particular ratios is much the same in the two cases. In fact, I should expect to obtain very similar results using estimates of $\frac{1}{2}$ and $\frac{1}{3}$ loudness, but would expect somewhat greater dispersions with estimates of, say, $\frac{3}{5}$ or $\frac{2}{7}$.

Whilst discussing this matter the authors refer to the case where a sound B is to a sound A as 2 : 1 and a third sound C "bears to B the relation of B to A", and they state that to assign the numeral 4 to C is "simply blundering." As one who has indeed done so, I am afraid that I am quite unrepentant. Surely this is only the question of interval again. If the difference between C and B is the same as the difference between B and A, then, as the authors say, the numeral 3 should be assigned to C. But if we accept

$$\phi(u, v) \equiv \frac{v}{u},$$

then when we say that $C : B :: B : A$ we mean that the ratio of C to B is the same as the ratio of B to A, in which case we must assign to C the number 4.

With regard to the facts set out in § 1 it has been usual, in the past, to consider that the loudness of a combination of sounds is independent of the relative phasing of the components. Recent work by Chapin and Firestone* however has cast doubt upon this. If the effect of phasing is significant, then fact (*e*) as stated is not sufficient and another variable must be introduced. Similarly relation (1) will not hold even if A and A' are pure tones of the same frequency.

The authors have provided such a valuable addition to the literature of this subject that it may seem ungracious to raise objection to minor points, but is not the term "mental estimates" sufficiently general to cover also the methods which the authors have classified as "equal relations"? I would like to suggest that the expression "mental estimates" be reserved as a general description of all methods involving estimates of sensations and that the particular process referred to in § 2 (i) be termed, say, "estimated interpolation."

Dr W. D. **Wright**. I believe that the problems which the authors are trying to investigate in this paper only exist because of two main misconceptions in their ideas. The first is that they do not yet appreciate the nature of a psychological entity. In the discussion on Dr Campbell's previous paper on the measurement of visual sensations, I had occasion to quote a statement of Mr Guild's and I feel that it is very necessary to repeat and emphasize it here. The statement was to the effect that "we cannot measure the magnitude of a sensation. We cannot experimentally isolate a unit of sensation from which to build up a quantitative scale of magnitude as required by all processes of measurement. This does not mean that we cannot make subjective estimates of the magnitude of a sensation; that in fact is the only means we have of obtaining quantitative information about sensations." If we wish to measure a length, we lay a calibrated scale of length alongside the unknown length and compare the two; if we wish to measure sensations, we must have a calibrated scale of sensation values before we can do so. Such a scale cannot be constructed and the attempt at measurement must be abandoned.

---

* E. K. Chapin and F. A. Firestone, *J. acoust. Soc. Amer.* **3**, 173 (Jan. 1933).

There is no evidence that this basic fact has been understood by the authors. One example may serve to illustrate this point. They contend that loudness is not additive because (p. 154) "ϵ can always be chosen so that if it is sounded simultaneously with a comparatively faint sound the combination is louder than that sound, but that if it is sounded simultaneously with a loud sound, the combination is no louder than that sound." This statement describes the effect, not of the addition of loudnesses, but of the addition of stimuli. The usual confusion between stimulus and sensation has occurred and the evidence proves nothing regarding the additiveness of loudness. The truth, as I see it, is that it is impossible to prove whether loudness is additive or not, because so long as we cannot isolate units of sensation it is impossible to carry out the experiments with the proper entities. I cannot myself conceive that it could be otherwise than additive, but that conviction is admittedly in no sense an experimental proof.

Although the authors entitle their paper "The measurement of loudness," they are not really concerned with such a measurement but are trying to find a comprehensive relation between stimulus magnitudes and sensation magnitudes, which is very different from the measurement of a sensation. But just as sensation-measurement is an impossibility owing to the nature of the quantity to be measured, so the derivation of a relation between stimulus and sensation is equally impossible, even if we could measure sensations, owing to the nature of the organ connecting the stimulus and the sensation. This is the second misconception of which the authors are victims. They seem to regard the sense organs as physical instruments in which, when you apply a certain stimulus at one end, you will always get the same response at the other. This, of course, is quite wrong. In the case of the eye, two differently coloured lights may be of equal brightness under one condition of the eye, but under another state of adaptation they may be of entirely different magnitudes. Evidently no stimulus-sensation relation can be postulated that will hold for both these conditions. It would be equivalent to using the same calibration curve for two very different instruments. Effects of this kind are possible for any of the sense organs, including the ear, and the facts must be faced. If they are faced, only one conclusion is possible, namely that no stimulus-sensation relation can be found which will be applicable to any but very limited variations of stimulation.

Dr L. F. Richardson. (i) What is a fact? It is customary in psychology to admit that phenomena peculiar to individual persons are facts. To demand universal validity would be unscientific. Once this general principle is conceded, it follows that mental estimates of loudness are facts, and therefore respectable. And we naturally pass on to consider the means and standard deviations of mental estimates made by numerous persons*. (ii) In a recent paper F. H. Gage shows that bisection of intervals of loudness leads to a remarkable misfit†. But Gage assumed that a fixed stimulus always produced the same sensation, thus neglecting successive contrast. Contrast may account for the misfit which Gage has found. (iii) Is it necessary to blame Fechner so severely for using an integral as an approximation to a sum?

Authors' reply. (1) Two facts, unknown to us previously, emerged in the discussion. We will state and examine them in a way concordant with the rest of our paper. The first is that on which the monaural-binaural scale depends. The fact here is that, given a sound B, another sound M can be found such that M heard monaurally is as loud as B heard binaurally. It is not a fact that to sound M must be attributed twice the loudness of sound B; that would be a fact only if binaural hearing were the addition, in the physical sense, of two monaural hearings. But a sound heard by one ear decreases the power of the other to hear very faint sounds; there is therefore evidence of a decrease in sensitivity accompanying the first step, which is fatal to additiveness. An analogy may help. If two similar bodies, each heated to 100° C., are dropped into water at 0° C., the rise of temperature is not twice that which occurs when one is dropped in. That is because the first body increases the heat-capacity and therefore decreases the sensitivity. There is therefore a definite reason for not calling the loudness of M twice that of B; if the monaural-binaural scale agrees with the 2 : 1 scale, that is evidence against the latter.

The second fact is that of Mr F. H. Gage. What he has shown is that relations, identified with equality because they are symmetrical, do not always (or usually) prove to be transitive. Equality in our discussion always implies transitiveness as well as symmetry. (For this reason we do not accept the definition of equality offered by Mr Floyd in his last paragraph.) The method of equal relations demands both qualities, as we pointed out. Accordingly within the range of Mr Gage's facts that method must be impossible; here we agree with him and Prof. Shaxby.

(2) For the rest we propose merely to underline and sometimes to expand statements made in our paper, taking them in the order in which they stand there. In addressing a Physical Society we felt justified in using "measurement" and its allied terms in the sense familiar to physicists, and ignoring entirely the very different sense in which they are used by psychologists. There seem to be two chief differences. First (if we understand Mr Floyd) psychological magnitudes are all what physicists call statistical magnitudes; the magnitudes we are concerned with

---

* L. F. Richardson and J. S. Ross, *J. gen. Psychol.* **3**, 288–360 (1930).

† F. H. Gage, *Proc. roy. Soc.* B, **116**, 103–119.

are not statistical, but simple. Secondly, if we understand Dr Wright, a "magnitude" can be "measured" psychologically, apart from all other magnitudes, whatever its qualities. In physics only additive magnitudes can be measured by themselves; all others are measured by means of relations between other magnitudes. Since loudness is not additive, it can only be measured in the second way. Accordingly establishing a relation between stimulus-magnitudes and sensation-magnitudes is not, for a physicist, a different thing from measuring loudness; it is the only possible way of measuring it.

As we said at the meeting, we do not believe that any part of our paper (except possibly § 4, which we do not defend seriously) has any relevance whatever to psychology. We address ourselves solely to physicists, who show by the discussion that they understand us except possibly in one particular: this is the significance of addition. The conditions stated in (1) are necessary, but not sufficient, criteria of additiveness. To state the sufficient criteria completely would take too long. Those who are inclined to doubt our statement that loudness is not additive should consider other cases of approximate additiveness, e.g. quantity of heat (mentioned above) or the output of a photoelectric rectifier cell.

(3) It is ungracious to quarrel with such a staunch supporter as Mr Guild; but we wish he would not speak of the association of *numbers* and magnitudes. The idea that numbers are associated with all magnitudes is the source of the fallacy which Dr Beatty upholds in his last paragraph. Only *A* magnitudes are associated with numbers; other magnitudes are associated only with *numerals*.

(4) We have never assumed responsibility for our "facts"; we have merely chosen those which, if they are facts, give the greatest chance of establishing a scale of loudness. If they are not facts, as some critics suggest, then the only conclusion is that the chance of establishing such a scale is even less than we estimated.

(5) In reply to a question put at the meeting, we define "intensity" by postulate; it is anything that makes (*e*) and (*f*) facts; both pressure and energy, measured in free space or in a telephone receiver, are intensities.

(6) We still refuse to define a fact formally. But it may help Dr L. F. Richardson if we point out that (in our sense) it is a fact that some people like blubber, but not a fact that blubber is delicious. Everyone agrees concerning the habits of the Eskimo, but not everyone shares their tastes.

(7) In discussing mental estimates we use "concordant" to mean "free from systematic divergence." In seeking agreement for technical purposes, it is permissible to ignore differences which, though systematic, are small. (Such differences are ignored, for example, in defining the colour-vision of a normal observer.) But if we are concerned to establish and use facts, small systematic differences are as bad as large ones. The facts that Dr Beatty and Mr Churcher themselves quote show systematic differences (*a*) between the estimates of different observers and (*b*) between the various scales which are said to agree. For some purposes the smallness of these differences may be important; for our purpose it is only their existence that matters.

Mr Constable's suggestion is interesting. The partial agreement between Mr

Churcher's various methods of obtaining a scale may arise from the use in all of them of the same arbitrary assumption concerning ϕ. If so, change of the assumption will alter all the scales in the same way and leave the agreement unchanged. For this reason even true agreement between several scales would not show that all are based on facts.

Our admission that systematic differences might be ignored for some purposes must not be taken as an admission that we regard those which Dr Beatty and Mr Churcher propose to ignore as actually negligible. On the contrary, in the branches of acoustics of which we have most experience, systematic discrepancies of 8 db. (a ratio of 6 : 1) would be very important indeed.

(8) We do not understand the relevance of Dr Beatty's theory of afferent impulses. If it is consistent with the facts that mental estimates differ and that loudness is not additive, it may be true but it adds nothing to the facts. If it is not consistent, it is simply false. Theories may explain facts; they cannot explain them away. If it should ever prove possible to count the afferent impulses, the theory might be converted into a new set of facts. But we cannot see how they could help. If the relation of the number of afferent impulses excited by a sound X to that excited by a sound Y were the same for all observers, the number may be a function of intensity, but it cannot possibly be a function of directly estimated loudness. If it were not the same for all observers, how should we be advanced in our search for relations valid for all observers?

(9) (Mr Cosens) No questions could fail to be leading questions, in our sense of the words, if the word "twice" were suggested at all to the observer. (Mr Davies) There is no blunder "in assigning the numeral 4 to C," unless at the same time "relation" is identified with arithmetical difference. If relation means arithmetical ratio, of course the assignment is right.

(10) We regard the physical errors of Fechnerism as even more disastrous than the mathematical; that is to say, the belief that all equations can be physically integrated, or that non-additive magnitudes can be added.

(11) We repeat that we do not regard a scale as necessarily useless because it is not based on facts. We accept fully Mr West's and Dr Rein's contention that certain scales are very useful in co-ordinating certain limited groups of observations. But we think they will find that their convenience arises merely from their providing a short name for a relatively complicated function of intensity; and that the observations could be described completely, though more cumbrously, without reference to anything but intensity.

(12) Finally we must deal with Mr Aldridge's plea. No scale would be satisfactory, in the very limited sense of this paper, unless it were wholly based on facts. If the facts on which a satisfactory scale could be based are not at present available, we could do what he asks us only by providing new facts. New facts can be provided only by discovery or by creation; for the present we are leaving these functions to others more competent than ourselves.

THE RECTIFYING PEAK VOLTMETER AS A STANDARD INSTRUMENT

By A. T. STARR, M.A., B.Sc., Faraday House, London

Addendum to paper previously published; received September 9, 1934.*

In § 6 of the paper as published the error due to the capacity of the screened leads from the standard condenser to the rectifier was discussed. The lead-capacity was shown to have a shunting effect whose value was calculated: it has, however, a very important effect on the error due to overbiasing, and discussion of this was unfortunately omitted.

Figure 13 of the paper shows the actual configuration of the rectifying peak voltmeter and figure 14 shows an exact equivalent. Appendix I shows that the overbiasing produces a constant error of e volts in the latter configuration, so that the error is

$$\left(e \div E\frac{C}{C+C_1}\right) \times 100 \text{ per cent,}$$

i.e.

$$\left(e\frac{C+C_1}{C} \div E\right) \times 100 \text{ per cent.}$$

The effect of the lead-capacity C_1 is thus to increase the overbias error from e to $e\,(C+C_1)/C$.

Dr L. G. Brazier discovered this effect experimentally. He used a sphere gap of $10\,\mu\mu$F. as condenser C, and the lead-capacity C_1 was $5000\,\mu\mu$F. The rectifying system was full-wave, the valves were LP. 4, the grid bias was 9 V. on each valve, and the overbias was in the neighbourhood of 70 V. Dr Brazier found that there was a constant error of 25 kV. r.m.s. with this arrangement. The theory given above shows that the overbias of 70 V. is multiplied by 501 because of the lead-capacity, so that the error is as follows:

$$\begin{aligned} 70 \times 501 &= 35 \text{ kV. (peak)} \\ &= 35/\sqrt{2} \text{ kV. r.m.s.} \\ &= 25 \text{ kV. r.m.s.} \end{aligned}$$

This agrees very well with the experimental values. It was in fact the experiment which led to the discovery of the omission in the theory, and the author wishes to thank Dr Brazier for having thus directed him to the omission.

The fact that the error due to overbiasing was constant at 25 kV. shows that the curvature of the valve characteristic had no appreciable effect.

In conclusion we can say that the lead-capacity may be as high as $20{,}000\,\mu\mu$F. so far as the shunting of the rectifier is concerned, but it must not be so large that $e\,(C+C_1)/C$ is an appreciable error. In practice the bias should be adjusted to be a little more than that required to suppress the zero current†. Then the bias should be intentionally increased by, say, 50 per cent or 75 per cent‡. If no change occurs in the rectified current, the error due to overbiasing is negligible.

---

* *Proc. phys. Soc.* **46**, 35 (1934).

† With LP. 4 valves a grid bias of 4·5 V. is required.

‡ With LP. 4 valves 1·5 or 3 V.

Magnetostriction oscillator
Undamped edg
Spongy rubber
Solid rubber

Figure A.

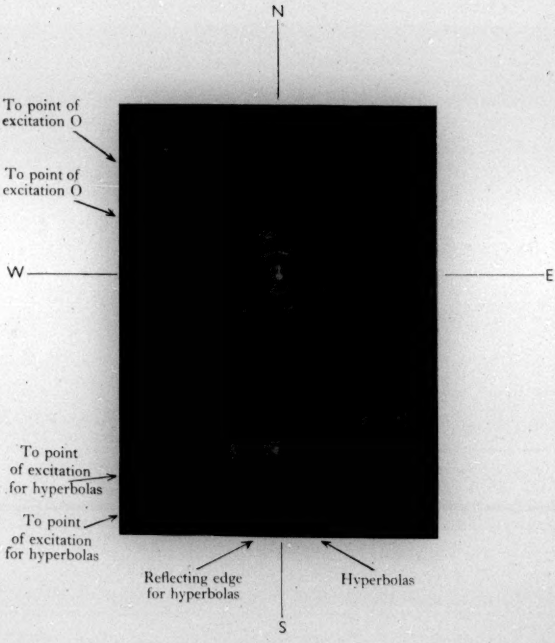

N
To point of excitation O
To point of excitation O
W
E
To point of excitation for hyperbolas
To point of excitation for hyperbolas
Reflecting edge for hyperbolas
Hyperbolas
S

Figure B.

DEMONSTRATIONS

"The velocity of sound in sheet materials." *Demonstration given by* A. B. Wood, D.Sc., F.Inst.P. and F. D. Smith, D.Sc., A.M.I.E.E. *November* 16, 1934.

A recent paper* on this subject describes a method of setting sheets of various materials into transverse vibration by bringing a vibrating nickel tube into contact with them. The reflection of these waves from the edges of the sheet produces a pattern of stationary waves which is made visible by sand scattered on the plate, after the manner of Chladni's figures. The separation of the nodes in the pattern is of the same order as the half-wave-length for transverse vibrations. The high frequency of vibration used by the authors results in a small wave-length and a pattern with a fine structure: in effect, a sheet of moderate size is large in comparison with the wave-length. The fact that the extent of the sheet includes a large number of wave-lengths has practical advantages. If the transverse vibration is subject to appreciable attenuation, as with rubber, cardboard, paper, ebonite and similar materials, the waves reflected from the remote edges of the sheet are practically extinguished: those reflected from the near edges produce simple interference patterns from which numerical values of the velocity of sound can readily be calculated.

To demonstrate the results described in the paper, sand patterns were formed on sheets of various materials by bringing a magnetostriction oscillator vibrating at 20,000 c./sec. into contact with the sheets. Complicated patterns of the usual Chladni type were formed on resonant materials, figure 1 of the paper: and rectangular hyperbolas were obtained on ebonite, figure 3 of the paper. The hyperbolas can be used to calculate the velocity of sound in the material.

In the demonstration a method of obtaining a simple pattern in a resonant material such as aluminium was shown, figure A. The aluminium sheet, of thickness 0·031 in., was damped by placing it between two sheets of spongy rubber except for a narrow strip along one edge. A light and uniform pressure was maintained on the damped portion of the sheet by a slab of solid rubber resting on the top. The undamped exposed strip was set into transverse vibration by bringing a magnetostriction oscillator into contact with it. The suppression of waves within the damped part of the sheet, without reflection at the damping edge, results in a simple pattern of parallel lines within the undamped part. Since these are spaced at half-wave-length intervals, the velocity of sound in aluminium can readily be found.

If the magnetostriction oscillator is applied to the sheet at an angle instead of normally, more complicated modes of vibration involving tangential movement of the sheet are excited. A curious pattern formed in this way on a sheet of keramot†, 0·031 in. thick, was shown, figure *B*. This material has appreciable internal damping and it is certain that the pattern is not formed by interference with edge-reflected waves. Similar patterns are also formed in other sheets, if edge-reflected waves are prevented by internal or external damping. The wave-length in the pattern is different from that of the simple transverse waves, as shown by the hyperbolas in the same photograph. It seems that the pattern is due to interference between vibrations of two types propagated outwards with different velocities from the point of excitation. One of these is probably a transverse vibration of the type considered in the paper and is excited by a force normal to the surface. The other is excited by a force tangential to the surface. The pattern is produced only when both components are present. Provided that the magnetostriction oscillator is applied normally to the sheet so as to excite simple transverse waves only, the anomalous patterns do not occur.

---

* Page 149 of this volume.

† A commercial insulating material.

In figure *B* the oscillator was placed in contact with the point marked *O* and inclined at an angle of about 45° to the line *SO* and at right angles to the line *EOW*. The normal component of force excited a transverse vibration radiating uniformly from *O*. The tangential component of force excited a tangential vibration radiating mainly in the directions *ON* and *OS*; these vibrations were in opposite phase, with the result that the interference patterns on the opposite sides of the line *EOW* were complementary, the nodes of one corresponding to the antinodes of the other. As there was no tangential force in the directions *OE* and *OW*, there were no interference patterns in their vicinity.

The transmission of high-frequency vibrations through a long thin wire mechanically connected to the magnetostriction oscillator was also demonstrated, the remote end of the wire producing transverse vibrations in sheets as in the cases mentioned above.

538.213

"A simple apparatus for measuring the pull between two magnetized surfaces."

Demonstration given by Prof. J. T. MacGregor-Morris and C. R. Stoner, B.Sc. (Eng.), A.M.I.E.E. *October* 19, 1934.

Introductory. A comprehensive examination of electrical-engineering and physics laboratories reveals the fact that experiments illustrating the relation between the magnetic flux and the pull between magnetized surfaces are seldom performed. The present investigation was started many years ago in the hope of meeting this need. The aim which the authors have had before them was to produce a robust and at the same time a fairly accurate piece of apparatus. A few laboratories have magnetic-pull permeammeters, but in some of these the apparatus is not used owing to its clumsiness in operation; and generally where any determined attempt has been made to correlate the pull with the flux, erratic results have been obtained and the value of the pull has been found to be too high, especially at low flux-densities. The apparatus about to be described was no exception as far as these matters were concerned, until a prolonged and concentrated effort had been made to master the difficulties.

The pull between two magnetized surfaces as related to the flux passing across the gap has been examined experimentally by several workers. The theoretical relationship was proved by Maxwell to be $P = B^2A/8\pi$, where P is the pull in dynes, B the flux-density in lines per cm², and A the cross-sectional area of the gap. Early experimenters obtained erratic results which were always higher than the values calculated from the formula, especially at low flux-densities. Later Threlfall discovered that this was due to the surfaces tending to tilt when parting; this increased the flux-density over a part of the surface, and although the area was decreased, yet because $P \propto B^2$ the pull might be greater under such conditions than with a lower B over a larger area. This effect is found to be particularly noticeable when B is small, because then the reduction of area brought about by the tilt does not produce saturation. Taylor-Jones, using guides and other refinements, obtained consistent experimental results which differed by only $\frac{1}{2}$ per cent from the calculated values between $B = 6000$ and $B = 10,000$.

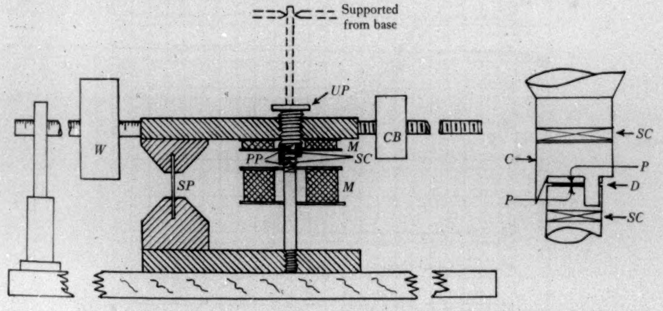


Figure 1. Magnetic-pull apparatus. *M, M*, magnetizing coils; *SC*, search coils; *PP*, surfaces which part; *UP*, upper pole-piece; *SP*, spring; *W*, weight; *CB*, counterpoise; *D*, non-magnetic disc; *C*, collar. [The left-hand drawing has been reduced to $\frac{1}{4}$, but the right-hand drawing has not been reduced.]

A number of workers have interested themselves in the problem of producing a permeammeter in which the value of B should be measured by the pull produced. Although, as we have seen, it is possible to produce such experimental conditions that agreement with theory is obtained, yet it is difficult to embody these conditions in a robust permeammeter. The important problem of flux-leakage has been almost entirely neglected in the permeammeters that have been developed, and either all the ampere-turns are assumed to be expended upon the specimen, or resort is made to a rather unsatisfactory calibration.

Description of the authors' apparatus. For their experimental investigations the authors have constructed the apparatus illustrated in figure 1. This has been designed to permit of comparisons between values of magnetic flux obtained from pull and search-coil measurements, and also an investigation has been begun into its possibilities as a permeammeter. Only the comparison work however will be discussed here.

It will be seen that the magnetic circuit is completed through a stalloy spring which is treated as a fulcrum; the dimensions are 15 × 1·9 × 0·06 cm. The effect of this spring when the apparatus is in use can be examined in figure 2, which shows the manner in which the pull between the surfaces falls off as the separation increases. This has been obtained from tests with different thicknesses of non-magnetic separators as described later. It will be seen that when the surfaces part, the decrease in attraction is so much more rapid than the decrease in the effective moment of the weight that the mechanical stiffness of the spring can be neglected.

The top coil is mounted upon the top beam and therefore any attraction between the current-carrying coils is added to the magnetic attraction. This force was both calculated and measured very approximately and found to be exceedingly small—less than the constraint produced by the spring. The attraction of the top beam by the lower coil was also found by experiment to be negligible.

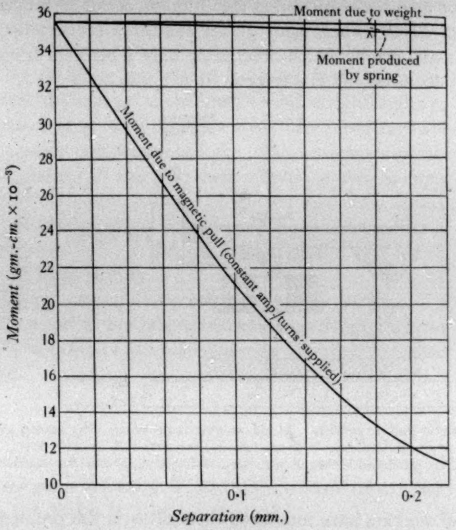


Figure 2.

Method of measuring flux-density and pull. Two similar search coils having an equal number of turns are wound one on each pole-piece. These are first opposed in a ballistic galvanometer circuit and the currents in the magnetizing coils are adjusted till minimum deflection of the galvanometer is obtained on reversal, this being less than 1 per cent of the deflection produced by either search coil alone. The search coil on the lower pole-piece is then used alone to measure the flux-linkages and hence B is obtained, the galvanometer being calibrated each time by means of an air-cored toroid of known dimensions.

The pull is measured by cautiously moving the weight along the beam until the magnetized surfaces part, noting the distance at which this occurs, and hence, by calculating moments, finding the pull at the gap. The current is then switched off, the gap is closed by sliding back the weight, and then the current is switched on in the reverse direction. In this way the magnetic circuit is kept in a cyclic state and readings can be repeated.

Results. In the apparatus in its original form comparison between B as measured by pull and the same quantity as measured by ballistic galvanometer showed erratic results, the pull-measurement being always too high; errors as great as 20 per cent were common at $B=6000$, and 10 per cent at $B=12,000$. Successive readings of the pull were as a rule consistent among themselves, but readings taken at different times usually showed quite different percentage errors. A systematic check of possible errors in the galvanometer-calibration etc. was made and failed to reveal any discrepancies.

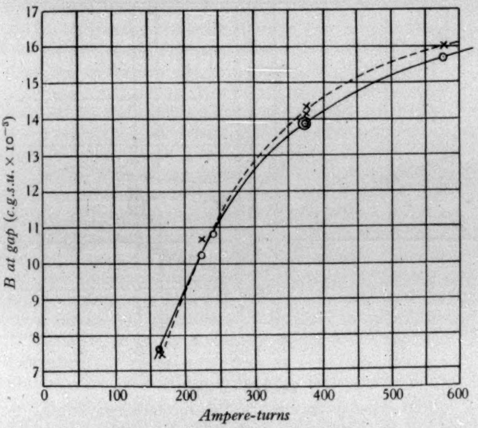


Figure 3. A typical set of results.

A study of the work of Threlfall and Taylor-Jones suggested that the irregularities were due to tilting on parting, and so a collar with three guiding claws was devised. A rod was also screwed into the top of the upper pole-piece and this slid through a ring-shaped guide supported from the base of the instrument as indicated by the dotted lines in figure 1. The screw thread on the upper pole-piece was opened out so that the guides could maintain a vertical movement of the pole-piece over a sufficient distance. The apparatus was now assembled while a magnetizing current was flowing. All constraints having been removed, the pole-piece was allowed to bed itself down magnetically into its correct position, and the guide was then locked in its place. A light was placed behind the gap so that the pole-piece and specimen could be seen to be in the correct position. The results obtained were disappointing, for they were as variable as before and errors were of the same magnitude.

It was then decided to insert a thin disc of non-magnetic material between the pole faces, so as to produce a small permanent gap with the object of reducing the cause for tilting. Now the errors were found to be greatly diminished, as figure 3 shows. Discs of various thicknesses were tried and figure 4 gives the percentage error for a number of discs of different thicknesses operating at various flux-densities. It will be seen that the use of a disc between 0·1 and 0·2 mm. thick brings the agreement between pull and ballistic-galvanometer measurements to within 5 per cent for values of B above 6000. The use of a thicker disc makes the pull-measurement deficient, owing in all probability to magnetic fringing.

Tests were then made with a 0·11 mm. disc, and on removal of the collar and guide the results were found to be almost identical with those obtained when the collar and guide were in use; it proved however desirable to retain the collar, as otherwise the disc occasionally worked out of the gap during the progress of a test.

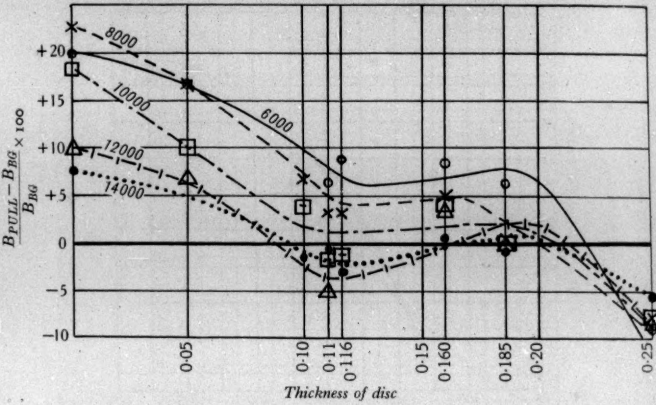


Figure 4. Curves of percentage error for different thicknesses of discs (in mm.).

These results appear to prove that the guide was insufficiently rigid to prevent the extremely small tilt which caused the error, and the insertion of a non-magnetic disc provides a much simpler solution.

It may be objected that it is bad policy to insert a definite air-gap in the magnetic circuit, but it should be remembered that no matter how perfectly the surfaces are worked up from the magnetic point of view there is always an equivalent air-gap, of small dimensions it is true, but still quite measurable in its magnetic effect and varying with the flux-density.

It appeared to be better, therefore, to insert a definite gap intentionally so that its effect would be more nearly constant at all flux-densities and at the same time the errors due to incorrect parting would be reduced to a negligible amount.

Conclusions. The apparatus has now been in the hands of students for some time and consistent results are obtained. It is believed to be more convenient and more accurate than any other apparatus which the authors have seen for the demonstration of the tractive force between magnetized surfaces.

REVIEWS OF BOOKS

Higher Mathematics for Engineers and Physicists, by I. S. SOKOLNIKOFF and E. S. SOKOLNIKOFF. Pp. xiii+482. (New York and London: McGraw-Hill.) 24*s*. net.

We know what to expect in a book entitled "Mathematics for Engineers"—a discussion of how many figures to retain in multiplying 19·34 by 14·17, a long disquisition on the differential equation $\ddot{y}+k^2\dot{y}+m^2y=\cos nx$, a bit about imaginaries, and all the rest of it. This is not one of the family. It is a book on *Higher* Mathematics for Engineers, and is an admirable introduction to its subject. For some reason (or rather, without reason), the authors apologize for addressing the book to physicists as well. Possibly conditions are different on the other side of the Atlantic, but certain it is that if every British physicist mastered this book, few of them would have wasted their time.

The authors plunge into their subject with a chapter on elliptic integrals, and then deal with the numerical solution of algebraic and transcendental equations. Horner's method is included, but one regrets to notice that the method of iteration is not mentioned. After a somewhat sketchy introduction to determinants and matrices, from which the multiplication law for matrices is omitted, we come to one of the best chapters in the book. This is on infinite series, and gives most of the common tests of convergence, deals clearly with uniform convergence, and even gives an example where Taylor's development of a function does not sum to the generating function. Then come chapters on partial differentiation, Fourier series, multiple integrals (with Jacobian introduced), line integrals and improper integrals, and ordinary differential equations. The chapter on the latter again is noteworthy for its thoroughness, dealing as it does with non-linear equations of the second order, and with the method of variation of parameters. It is useful to find in the chapter on line integrals a proper investigation of the conditions under which the value of the integral is independent of the path.

The chapters on partial differential equations, vector analysis, and probability and curve-fitting are less distinguished, whilst the final chapter, on conformal representation, is again very good. This chapter is extracted from a lecture by another author, but one feels after reading the rest of the book that the Sokolnikoffs could have written one at least as good, and the rather noticeable change of style would not then have occurred.

The book is well stocked with examples, mostly rather easy, and the publishers are to be commended on their foresight in printing the words "first edition" on the title-page.

J.H.A.

Mathematical Problems of Radiative Equilibrium, by EBERHARD HOPF. Pp. vi+105. (Cambridge Tracts in Mathematics and Mathematical Physics, No. 31. Camb. Univ. Press.) 6*s*. net.

This tract treats of problems connected with a certain integral equation which occurs on the theory of the outer layers of a star. To quote the preface: "The astrophysicists mostly contented themselves with an approximate solution of these problems. Owing to a certain inherent beauty, however, they aroused also the interest of the rigorous mathematician. It is the purpose of this tract to attempt a coherent representation of all that has been achieved in the direction of a rigorous solution of those standard problems." For the purpose in view, the tract is of high excellence.

In reading it, I have to struggle with a feeling that it is rather sacrilegious to treat a star in this way. Of course, it is not an actual star that is butchered to make a pure mathematician's holiday; it is a "gray" star invented by Schwarzschild, or a purely scattering star invented by Schuster. I fear the reader may not realize that, although the astrophysicist does not treat these problems rigorously, he treats them more rigorously than an actual star is likely to do. Certainly the rigorous mathematicians have secured one triumph. There had been a kind of competition to give the best approximation to the ratio of the boundary temperature to the effective temperature in a gray star. The fourth power of this ratio was first given by Schwarzschild as 0·5; there were four rival later approximations ranging between 0·42 and 0·44. Hopf and Bronstein, by difficult mathematics which occupies many pages in this tract, found that the exact value was $\sqrt{3}/4$.

It may not be out of place to recall that Sir Arthur Schuster, whom we have lately lost, was the pioneer on the astrophysical side of this subject, and his paper "Radiation through a Foggy Atmosphere" (1905) is continually referred to by modern writers as a classic.

A.S.E.

Theoretical Physics, by G. Joos. Translated from the first German edition by I. M. Freeman. Pp. xxiv + 748. (London: Blackie and Son, Ltd.) Price 25*s*.

The first German edition of this book, published in 1932, had an immediate success, and the second edition is already announced. A new chapter on nuclear physics, which is being added to the second German edition, is incorporated in the English translation.

The scope and general nature of the book are already well known to many English users, and the briefest description may suffice here. The text opens with a seventy-page mathematical introduction. This is followed by six sections dealing in turn with mechanics (including relativity mechanics), the field theory of electro-magnetic and optical phenomena, the atomistic nature of electrical phenomena, the phenomenological part of heat theory, the statistical part of heat theory (including quantum statistics), and finally nuclear, atomic and molecular structure and spectral theory.

There are, of course, omissions, but the material has been very well selected, and the treatment is admirably suited to the needs of the average serious student of physics. Here and there the text is supplemented by examples, to which worked solutions are supplied. It is not to be expected that a book of this size should give an adequate knowledge of advanced physics, but Prof. Joos has written a text which may serve as an excellent introduction to more detailed study or as a valuable synopsis of knowledge properly acquired in a less concentrated form.

The publishers have hit upon the singularly happy idea of employing a translator with a knowledge of the subject *and* of the two necessary languages; in brief, the translation is good, and we only rarely encounter passages which might have been clearer if the translator had been willing to take greater liberties with the original phrasing. The book is well produced, and is remarkably good value at the price.

The one adverse criticism, which applies to any sound translation of a good *Lehrbuch*, is that the translation plays too much into the hands of those science students who are already only too content to remain bi-, uni-, or even semi-lingual. This is a pity.

H.R.R.

Atomic Structure and Spectral Lines, Vol. I, by A. Sommerfeld. Translated from the fifth German edition by H. L. Brose, M.A., D.Phil., D.Sc. Pp. xi + 675 with 151 diagrams. (Third English edition. London: Methuen and Co. Ltd.) 35*s*. net.

In the latest German edition of Sommerfeld's classic work the subject matter is divided into two volumes, volume I, which appeared early in 1932, being a revision of the fourth (single-volume) German edition of the original *Atombau und Spektrallinien*, and volume II an elaboration of the supplementary work *Wellenmechanische Ergänzungsband* of 1928. Accordingly the new English edition is also to appear in two volumes, the first of which is now before us. Compared with the two previous English editions, this volume is in part an abbreviation and in part an extension; its scope will therefore be pretty well known to many readers, and need not be outlined here. Since, as Prof. Sommerfeld himself remarks in his preface, it is possible to understand the new quantum theory only by building it up from the old, the present volume deals both with the fundamental experimental material and also with the conception of orbits as a means of introducing the quantum numbers and as models for subsequent wave-mechanical calculations. The results established by the older theory are given throughout in the forms appropriate to the newer mechanics, the later proofs being reserved for the next volume. The methods of Hamiltonian mechanics, which previously appeared in the Mathematical Appendix, have now been included in the text, whilst certain other matter has been transferred from the text to the appendix. The treatment of multiplet structures has been revised and extended by the inclusion of Pauli's Principle and electron spin. In view of the extent and importance of the recent developments of the study of band spectra, the treatment of this subject in the forty-page chapter to which Prof. A. Kratzer and Dr K. Bechert have contributed seems rather inadequate. Prof. Sommerfeld, it is true, points out in his preface that this short chapter makes no claim to be in any way complete, and contains only what is essential for wave-mechanics.

It is pleasant to see that the figures in the new edition are supplied with captions which were entirely lacking in the earlier editions. Prof. Brose states in his preface that the author wished the translation not to be too literal, slight modifications being left to the translator's discretion. Comparison of certain passages in the first edition with corresponding ones in the new shows that his exercise of this privilege has improved the work. Occasionally one wishes he had gone even farther in this direction, and abandoned some symbols, words, and phrases with which the German reader may be at home but the English reader is not. We are not, for instance, in the habit of using the words "partial band," "R-band," "P-band," "edge" and "disturbance" (Chapter IX) for "band," "R-branch," "P-branch," "head" and "perturbation" respectively; and as a symbol for moment of inertia we prefer I to $\mathcal{J}$ in a text in which the latter is also used for a quantum number. And why retain both j and $\mathcal{J}$ for one and the same quantum number in the same chapter? We could also wish that it were permissible for a translator to jettison the misnomer "Bergmann series" (p. 352), since the term symbol mb of former editions is now replaced by nF; even if the late Dr Hicks's designation "fundamental series" had not been adopted, there would have been little difficulty and much justice in coupling another investigator's name with these series, for they were discovered by Saunders and by A. Fowler before Bergmann, working in Paschen's laboratory, investigated them, and their significance was first clearly pointed out by Runge.

This English edition contains two sections which are not in the German editions but have deservedly received the approval of Prof. Sommerfeld. One of them is an excellent six-page account of hyper-fine structure and nuclear moments by Dr E. Gwynne Jones; it is well written and admirably suited to the main scheme of the work. The other is a very welcome addendum of five pages by Prof. Brose describing briefly the advances which have been made in atomic physics in the interval between the appearances of the last German edition and the present volume; the subjects of this addendum include investigations of cosmic radiation and nuclear transformations, and the discoveries of the positron, the neutron and the isotopic constitution of hydrogen.

In so large a volume minor slips, especially in the matter of names and initials, are almost unavoidable; only a few so far have been noticed by the reviewer. "A. A. Millikan" (p. 15), "Giaque" (p. 142), "J. W. Aston" (p. 171), "square brackets" (p. 551), "nucleii" (p. 551), "Ann Arbor University" (p. 562), "Lewin" (pp. 566–7), "H.COOH" (p. 586), and "Birheland" (p. 598). The second sentence of the last paragraph on p. 559

is meaningless, owing, as one finds on looking up the corresponding sentence in the first edition, to the omission of several words. On p. 561 "equidistant sequences of lines" should be "sequences of equidistant lines."

English and American students of atomic physics will feel more indebted than ever to Prof. Brose, and will welcome this new edition even more cordially than the old.

W.J.

The Kinetic Theory of Gases (*Some Modern Aspects*), by Prof. M. KNUDSEN. Pp. 64. (London: Methuen and Co., Ltd.) Price 2*s.* 6*d.*

This book practically reproduces the lectures given by Prof. Knudsen at King's College, London, in 1933. It deals almost exclusively with the author's well-known investigations of phenomena occurring at low gas pressures. As a clear and concise first-hand account of these important researches it is assured of a good reception. The beginner should, however, be warned that "Some Modern Aspects" is the operative phrase in the description of the text—he will find no summary of general kinetic theory. H.R.R.

Geometrische Elektronenoptik, by E. BRÜCHE and O. SCHERZER. (Berlin: Springer.) RM 26; cloth-bound RM 28.40.

Although the general phenomena of the effects of electric and magnetic fields on electric discharges have been known for many years, it was only in 1926 that H. Bush showed theoretically that if such fields possessed an axial symmetry they could be used to focus the paths of electrons derived from a single point of origin. The full significance of this work was not at once realized and, broadly speaking, it was not followed up for some four years; but in the period from 1930 to the present time a great deal of work has been done which is ably summarized in the book under review. Not only have the practical possibilities been studied in the period mentioned, but the realization of the wave-directed characteristics of electrons in motion has enabled Glaser and others to apply Hamiltonian methods to the discussion of the theory.

On the practical side, the new methods have proved of the greatest value for the study of thermionic emission from surfaces, and have had valuable applications in the design of cathode-ray oscillograph tubes. It has also proved possible to obtain real images of objects irradiated by electron streams; these will bear a magnification of 8000 or more, even with the employment of magnetic fields for which no attempt has been made to eliminate the aberrations which are known to be present. Theory indicates an ultimately possible resolving-power far out-reaching that of optical microscopy, and it is by no means impossible to deal with biological objects by special methods. The book presents a very complete picture of such practical developments up to the present date, and gives useful chapters on the theory, although the theoretical presentation makes a less coherent whole. The wave-mechanical theory, in spite of its mathematical beauty, proves over-strenuous, and more progress can be made for practical purposes by more straightforward methods.

Since there is such a complete analogy with Gaussian Optics in many respects, we may, perhaps, ask in passing why more attention is not given to the symmetrical form of the paraxial equations. A refractive index appears, for example, as proportional to the square root of a potential, and it would seem to be generally useful to quote the conjugate relation in such a form as

$$\frac{\sqrt{\varphi'}}{l'}-\frac{\sqrt{\varphi}}{l}=\frac{1}{8}\int_{-\infty}^{+\infty}\varphi_z^{-\frac{3}{2}}\,(\partial\varphi_z/dz)^2\,dz=\frac{\sqrt{\varphi'}}{f'},$$

where l and l' are conjugate distances. The corresponding Gaussian equation is

$$\frac{n'}{l'}-\frac{n}{l}=\frac{n'}{f'}=-\frac{n}{f},$$

if a well-known sign convention is used. Such equations convey far more information than those usual in the book, which merely give an expression of the "focal length" as the result of a theoretical discussion.

In the discussion of such a large range of theoretical and practical matters a good deal is bound to lie outside the practical experience of any author, and it will not be surprising if such a book presents, on the whole, too easy an appearance to the whole subject. The phenomena of the discharge tube are not by any means completely mastered, and although what appear to be results of surprising promise in "microscopy" have already been achieved, there can be few subjects in which real advances will require more patience, skill, and technical resource. Consider some of the factors involved in the present image-production; the very small angular apertures of the beams which have so far been found practicable for use, the heterogeneous character of the electron-energies, the aberrations of the lenses, only measurable with difficulty, and the various difficulties of vacuum work where high vacua must be maintained. It will be seen that the superficial reading of a text-book may easily produce a mental picture glowing in over-rosy hues. On the other hand the book does present a picture of a technical method which, although difficult, reveals no such unsurmountable barrier as that produced by the wave-length of light in optical microscopy; no ultimate barrier, that is, in dealing with smaller and smaller objects until the region of the phantom electron wave-length is reached, in the neighbourhood of 10^{-8} cm. or less. If technical methods of dealing with biological objects are successfully developed, the possibilities of advance are incalculable, and methods of microscopical examination of bodies far beyond resolving-power even of the ultra-violet microscope seem to be within the range of expectation. L.C.M.

Diffraction of X-rays and electrons by amorphous solids, liquids and gases, by J. T. Randall, M.Sc. Pp. xii + 290. (London: Chapman and Hall, Ltd.) 21*s*. net.

This book appears at an opportune moment. The discussions on the solid state at the recent conference showed that the centre of interest is shifting from a consideration of the structure of regular crystals to that of the irregularities which disturb the lattice and of the various non-crystalline and quasi-crystalline structures with which this book deals. In particular the problem of the transition from a liquid or a glassy state to perfect crystalline arrangement seems likely to occupy an increasing amount of attention in the next few years.

The author has performed a valuable task in collecting together the very large amount of scattered work which bears on these problems and presenting it in a clear and logical order. His work will save a great deal of time and trouble to all working in this field and enable them to see more clearly the directions in which to advance.

Apart from its purely scientific interest, the subject is of considerable industrial interest, especially as regards the study of glasses and organic fibres, so that it is appropriate that the author should be a member of one of the great commercial research institutions. The book is extremely up-to-date, and it is interesting to see the stress laid on electron-diffraction as an experimental method which has already proved its usefulness in this field and seems likely to have a rapid development. On the theoretical side the chief results are given with an indication of the way in which they have been obtained, but as a rule without detailed analysis. G.P.T.

Thermionic Emission, by Arnold L. Reimann. Pp. xi + 324. (London: Chapman and Hall, Ltd.) 21*s*. net.

In a volume of 324 pages the author gives an admirable survey of a subject which has received a vast amount of attention in recent years, particularly in the industrial research laboratories of manufacturers of thermionic valves, and it may be mentioned that the author is on the staff of the G.E.C. laboratories.

In the last decade big additions have been made to our theoretical and practical knowledge of thermionic phenomena. The classical theory has been replaced by the quantum statistics in the theory of electrons in metals and in a vacuum, and by the application of the wave-mechanical theory of the transmission of electrons through potential barriers. On the practical side the employment of modern high-vacuum technique has resulted in a marked increase in the quantity and accuracy of the data available.

The volume is primarily concerned with thermal electron-emission and the principal emitters are the metals and certain other electronic conductors. Thirty-three years have elapsed since O. W. Richardson derived his emission formula to fit the experimental results for the emission of electrons from metals. In his theory the velocity-distribution of the electrons available for emission was assumed to be the Maxwellian. Later this theory was involved in so many difficulties that it had to be abandoned and has been replaced by the theory of electrons in metals originally put forward by Pauli and Sommerfeld and since modified and extended by Bloch and others. This theory assumes the existence of effectively free electrons, whose velocity-distribution, instead of being Maxwellian, is governed by the new quantum statistical principles of Fermi and Dirac.

It has been shown by Nordheim and Sommerfeld that the substitution in Richardson's derivation of his original formula of the Fermi-Dirac distribution for the Maxwellian distribution leads to an emission formula concerning whose essential validity there can be no reasonable doubt. It is interesting to note in this connection that H. A. Wilson deduced an emission formula by thermodynamic reasoning without any assumption as to the actual mechanism of the emission, and this theory was further developed by Wilson and Richardson. Later developments of the thermodynamical theory are due to M. V. Laue, W. Schottky and S. Dushman. In its final form the formula derived thermodynamically agrees perfectly with that obtained by Nordheim and Sommerfeld by the quantum statistical method referred to above.

Hence it will be apparent that the study of thermionics touches directly on the electrical constitution of conductors and that the thermal emission of electrons is related to other electron emission phenomena such as the photoelectric effect. In the present volume the theory is presented in a simple form, and where it has not been possible to discuss in any detail such branches of physics as quantum statistics and the modern electron theory of metals, the author has tried to show by what processes of reasoning the results have been arrived at.

The book should prove of great value to those interested in the various practical applications of thermionic phenomena, and the author is to be warmly congratulated on the results of his labours in the production of the volume under review. E.G.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 47, PART 2 March 1, 1935 No. 259

538.22:546.711

THE MAGNETIC PROPERTIES OF AMORPHOUS MANGANESE

ERRATUM

"An X-ray study of the effect of heat on the structure of sputtered films of gold", by S. RAMA SWAMY, B.Sc., Ph.D., *Proc. phys. Soc.* **46**, 739 (1934).

In line 4 on page 742

for $\tan\phi = L_0/(d \sin\theta)$ *read* $\tan\phi = L_0/(d \sin 2\theta)$

contain about 3·4 parts of ferromagnetic impurity per 1000, and measured its susceptibility over a wide range of temperature, the value at room temperature being $9{\cdot}70 \times 10^{-6}$ e.m.u. per gram. A little later, Ishiwara$^{(8)}$, using a similar sample, obtained $9{\cdot}66 \times 10^{-6}$, and suggested that certain traces of ferromagnetism might be attributed to the combination of manganese with nitrogen.

Hadfield, Chéveneau and Géneau$^{(5)}$ prepared manganese from an amalgam and cast it in dry hydrogen. The grey powder which they obtained was slightly ferromagnetic with a susceptibility of $11{\cdot}2 \times 10^{-6}$, the traces of ferromagnetism being attributed to combination with hydrogen. Freese$^{(3)}$, however, found that manganese prepared in this way was non-ferromagnetic, and that specimens prepared by cathode disintegration in hydrogen showed no traces of ferromagnetism. Kapitza$^{(9)}$ found that a very pure specimen of manganese prepared by distillation and melted *in vacuo* by Miss Gayler$^{(4)}$ was perfectly paramagnetic with a susceptibility of $9{\cdot}66 \times 10^{-6}$ at room temperature. Miss Wheeler$^{(20)}$ prepared samples of α-manganese by distillation and obtained a sample of β-manganese by quenching an α specimen

In the last decade big additions have been made to our theoretical and practical knowledge of thermionic phenomena. The classical theory has been replaced by the quantum statistics in the theory of electrons in metals and in a vacuum, and by the application of the wave-mechanical theory of the transmission of electrons through potential barriers. On the practical side the employment of modern high-vacuum technique has resulted in a marked increase in the quantity and accuracy of the data available.

The volume is primarily concerned with thermal electron-emission and the principal emitters are the metals and certain other electronic conductors. Thirty-three years have elapsed since O. W. Richardson derived his emission formula to fit the experimental results for the emission of electrons from metals. In his theory the velocity-distribution of the electrons available for emission was assumed to be the Maxwellian. Later this theory was involved in so many difficulties that it had to be abandoned and has been replaced by

cations of thermionic phenomena, and the author is to be warmly congratulated on the results of his labours in the production of the volume under review. E.G.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 47, PART 2 March 1, 1935 No. 259

538.22:546.711

THE MAGNETIC PROPERTIES OF AMORPHOUS MANGANESE

BY L. F. BATES, D.SC., PH.D., F.INST.P., Reader in Physics,

AND

D. V. REDDI PANTULU, B.SC., University College, London

Received October 13, 1934. *Read in title December* 7, 1934.

ABSTRACT. A description is given of the preparation *in vacuo* of pure amorphous manganese. It is found to be paramagnetic, without trace of ferromagnetism, and to obey the Curie-Weiss law, $\chi=2\cdot174\times10^{-2}/T+1540$, over the range of temperature 90 to 600° K., the experimental value of χ at 20° C. being $11\cdot80\times10^{-6}$.

---

§ 1. INTRODUCTION

IN recent years the magnetic properties of manganese have been investigated by many workers, for it is known to exist in several well-defined states and to combine readily with non-magnetic elements to form ferromagnetic substances. In 1912, Honda(6) used a specimen of fused manganese obtained from Kahlbaum, stated to contain about 3·4 parts of ferromagnetic impurity per 1000, and measured its susceptibility over a wide range of temperature, the value at room temperature being $9\cdot70\times10^{-6}$ e.m.u. per gram. A little later, Ishiwara(8), using a similar sample, obtained $9\cdot66\times10^{-6}$, and suggested that certain traces of ferromagnetism might be attributed to the combination of manganese with nitrogen.

Hadfield, Chéveneau and Géneau(5) prepared manganese from an amalgam and cast it in dry hydrogen. The grey powder which they obtained was slightly ferromagnetic with a susceptibility of $11\cdot2\times10^{-6}$, the traces of ferromagnetism being attributed to combination with hydrogen. Freese(3), however, found that manganese prepared in this way was non-ferromagnetic, and that specimens prepared by cathode disintegration in hydrogen showed no traces of ferromagnetism. Kapitza(9) found that a very pure specimen of manganese prepared by distillation and melted *in vacuo* by Miss Gayler(4) was perfectly paramagnetic with a susceptibility of $9\cdot66\times10^{-6}$ at room temperature. Miss Wheeler(20) prepared samples of α-manganese by distillation and obtained a sample of β-manganese by quenching an α specimen at 1000° C. in water. She found susceptibilities of 9·60 and $8{\cdot}80 \times 10^{-6}$ for the α and β specimens respectively, with no traces of ferromagnetism. It must also be recorded that Shimizu(14) obtained specimens by distillation and followed the magnetic changes associated with the α to β and β to γ transitions, but his values for the susceptibilities are much lower than those recorded above, and the two magnetic transition points recorded by him do not appear to correspond with any of the four transition temperatures recorded by Miss Gayler.

It is generally stated that the susceptibility of manganese appears to be independent of the temperature over a wide range. Thus, Owen(12) used a fused Kahlbaum specimen and found a value of $8{\cdot}93 \times 10^{-6}$ which was practically independent of temperature over many hundreds of degrees. This value is, however, rather low compared with Miss Wheeler's, assuming that it refers to unquenched manganese. Again, Honda and Soné(7) made measurements with similar specimens and also found the susceptibility constant over a wide range. It does not appear to be profitable, however, to consider their results in detail, for their specimens contained considerable quantities of iron and appreciable amounts of aluminium, copper and silicon, and certain sudden changes in susceptibility and the temperature hysteresis phenomena exhibited by these specimens may not be attributed to pure manganese. It therefore seems that the statements concerning the constant susceptibility of manganese are open to question. As no record of the preparation of pure amorphous manganese from an amalgam heated *in vacuo* could be found, and its thermomagnetic properties were unknown, the following preparation and investigation were made.

§ 2. PREPARATION OF SPECIMENS

Manganese amalgam was prepared by electrolysis in the following manner. A saturated solution of manganese chloride was placed in a large glazed earthenware trough and crystals were added during the electrolysis to keep the solution saturated. The anode consisted of a platinum strip placed inside a porous pot containing saturated manganese sulphate solution. The cathode was a pool of mercury contained in a flat dish. During electrolysis the reactions at the anode were vigorous and could become unpleasant, but the latter tendency was counteracted by periodically transferring the platinum strip to a porous pot filled with fresh solution. A stream of manganese sulphate could, of course, be used in large-scale manufacture.

Usually, a current of six amperes was passed for thirty minutes. The method employed in stirring the mercury was important, for when it was vigorously stirred by a rotating glass vane the amalgam did not appear to be at all rich in manganese. Careful stirring with a glass stirrer, such that the portions of the amalgam richer in manganese were systematically pushed beneath the surface of the pool, gave very satisfactory yields.

The amalgam was quickly washed in water and squeezed in a linen bag. The heavy mass thus obtained was rapidly forced into the pyrex glass bulb *A* of figure 1 through the tube *a*, which was then sealed. The apparatus of figure 1 was placed inside a furnace up to the region *b*, and exhausted through drying tubes by a Hyvac pump. When all traces of water had been removed the temperature was raised slowly to 370° C., when the mercury vapour driven off was condensed in a water-cooled bulb. After the bulb A had been maintained at 370° C. for a sufficient time a mercury-vapour pump and vapour-trap were used to remove the last traces of mercury from the manganese. The apparatus was sealed at the constriction e and allowed to cool slowly.

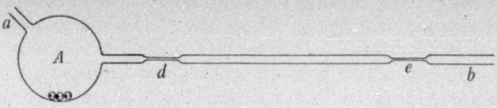


Figure 1.

The porous black lump of manganese was now powdered and forced into the tube de, which was of uniform cross-section of diameter 0·5 cm. Inside the bulb A were three pyrex beads, and on shaking of the bulb these easily smashed the manganese to a fine powder. The powder was shaken into the tube, the beads again being used to force it through the constriction d. Finally, when practically all the manganese had been forced into de, the tube was sealed at d, and glass hooks were formed at d and e as shown in figure 2. The tube de now contained a specimen of amorphous manganese prepared and sealed *in vacuo*. Its purity undoubtedly depended to some extent on the rapidity with which the amalgam was transferred from the electrolyte trough to the bulb A.

§ 3. MEASUREMENT OF MAGNETIC SUSCEPTIBILITY

The measurement of magnetic susceptibility was made by the Gouy method, the tube de being suspended vertically from the arm of a sensitive balance with its lower end in a uniform field between the flat pole pieces of an electromagnet, as shown in figure 2. A tube cut from the same piece of pyrex was evacuated, sealed and attached close to the lower end of de, in order to compensate for the magnetic effects of the tube de and the air displaced by it. Therefore, except for a small effect due to slight dissimilarities between the two tubes, the magnetic field acting on the system provided a downward force mg equal to $\frac{1}{2}k\alpha\,(H^2-H_0^2)$ dynes, where k is the susceptibility per cm.^{3} of the manganese, α the area of internal cross-section of the tube, and H and H_0 the strengths of the field at the lower and upper ends of de, respectively. In our experiments H_0^2 and the small effect just mentioned were found negligible. H was measured by a calibrated fluximeter and search coil.

mg

H

In the high-temperature measurements de was suspended from a copper wire inside the furnace, as shown in figure 2. The furnace consisted of a thick-walled copper tube suitably wound and lagged. The temperature in the neighbourhood of the lower end of de was recorded by a mercury thermometer inserted through f; it was calibrated *in situ* against a series of standard thermometers placed vertically in turn inside the furnace. In the low-temperature measurements the tubes were suspended by a thread inside a copper tube surrounded by liquid oxygen or carbon dioxide and alcohol mixture contained in a Dewar flask. In the latter case the temperature was measured by a standard pentane thermometer.

After these measurements had been made, *de* was weighed and opened, and the manganese was removed. The area of internal cross-section and the volume previously occupied by manganese were found by suitable weighings. The susceptibility per gram could now be calculated. These operations meant the loss of the pure manganese, for we were unable satisfactorily to transfer the intensely pyrophoric material from one vessel to another. The method of poisoning with hydrogen which we tried to adopt in accordance with published information was found useless. Indeed, very little hydrogen appeared to be absorbed by the manganese. The above magnetic results were also checked against a standardized solution of nickel chloride of known susceptibility, kindly supplied by Professor Sugden.

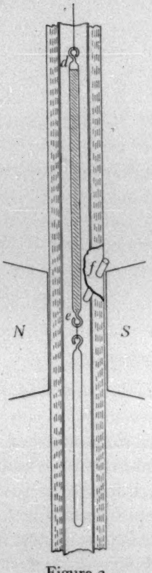


Figure 2.

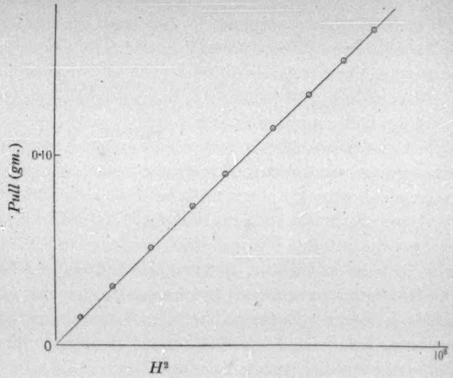


Figure 3 (*a*). Specimen No. 1.

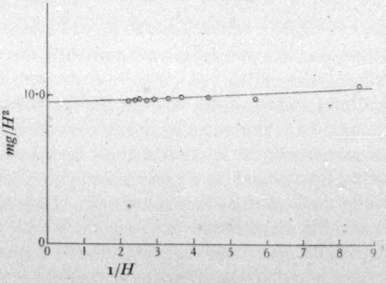


Figure 3 (*b*).

§4. EXPERIMENTAL RESULTS

The most complete investigations were made with two chosen specimens of amorphous manganese. The first specimen, I, was made before the best arrangements had been effected for the rapid handling of the amalgam, so that it was in contact with air for a longer period than was desirable. Accordingly, it exhibited traces of ferromagnetism as shown by figures 3 (*a*) and 3 (*b*), where the downward force mg is plotted against H^2, and mg/H^2 against $1/H$, respectively. For small values of H^2 the curve of figure 3 (*a*) is not linear. Figure 3 (*b*) shows that Honda's method may be employed to eliminate the effect of the ferromagnetic impurity, provided that the values of H^2 are not so low that the measurements are liable to be inaccurate.

The second specimen, II, was made under very satisfactory conditions, and showed no traces of ferromagnetism, the curve corresponding to figure 3 (*a*) being strictly linear. Mr H. Terrey kindly analysed this specimen and found it entirely free of mercury and other metals. We therefore conclude that amorphous manganese properly prepared *in vacuo* is definitely non-ferromagnetic, and the traces of ferromagnetism exhibited by specimen I are attributed to the action of nitrogen; we propose to study this action in more detail at a later date.

The way in which the susceptibility of specimen II varied with temperature over the range −183 to 340° C. is shown in figure 4, where the reciprocal of the susceptibility per gram is plotted against the temperature. Figure 4 shows that over a considerable range the value of the susceptibility per gram is given by the Curie-Weiss law

$$\chi = C/(T+\Delta) = 2{\cdot}174\times10^{-2}/(T+1540),$$ Δ, T

from which it follows that the effective magnetic moment of a manganese atom is 3·09 Bohr magnetons. The actual susceptibility per gram at 20° was $11{\cdot}80\times10^{-6}$. All values in figure 4 have been corrected for the diamagnetism of the manganese atom, which was taken as 0·1 unit per gram(16).

Perhaps a more adequate conception of the rate of change of χ with T may be obtained by reference to figure 5, where the uncorrected experimental values of the susceptibility are plotted against the corresponding absolute temperatures.

In order to compare the paramagnetic properties of the two specimens, the effects of the ferromagnetic impurity in specimen I were eliminated by Honda's method; see figure 3 (*b*). It was considered unnecessary to apply the more complete form of correction given by Vogt(17, 18). The susceptibility χ_∞ in an infinitely large χ_∞
field is assumed to be related to the susceptibility χ_H in a field H by the equation χ_H
$\chi_\infty = \chi_H - \sigma/H$, where σ is the saturation magnetization of the ferromagnetic impurity. The values of χ_∞ were therefore obtained at chosen temperatures −183, 20, 166 and 299° C., from the corresponding values of χ_H for two suitable values of H. The graph of $1/\chi_\infty$ against T was strictly linear over the range −183 to 300° C. with a value 1541° of Δ. The value of the susceptibility at room-temperature corrected for diamagnetism of the manganese atom was $11{\cdot}83\times10^{-6}$. The agreement with the corresponding values of specimen II is very striking.

Two further points of interest appear to merit record in connexion with specimen I. First, the mass of powder per cm.^{3} inside the tube was some 20 per cent greater than in the case of specimen II. Secondly, specimen I was far less pyrophoric, and could be transferred from one tube to another with comparative ease in an atmosphere of hydrogen. These points deserve more extensive examination than we have so far been able to give them.

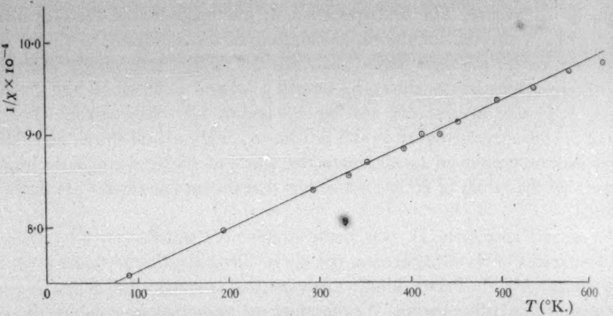


Figure 4.

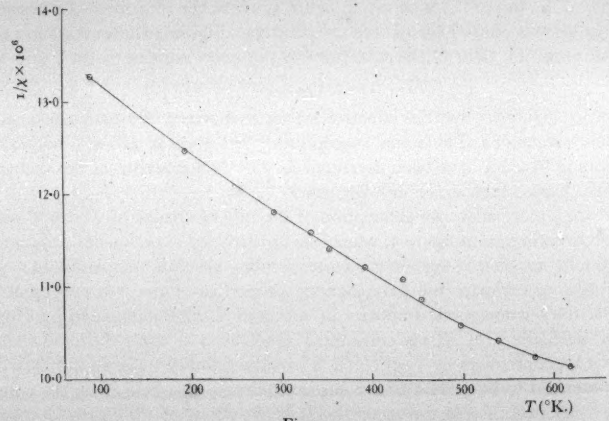


Figure 5.

§ 5. DISCUSSION OF RESULTS

The value obtained for the susceptibility at room temperature compares favourably with the corrected value $11\cdot3\times10^{-6}$ obtained by Hadfield and his collaborators, in view of the differences in the preparations. There is, however, a considerable difference between our value and those of Kapitza and Miss Wheeler, which must be due to differences in crystalline structure. It would thus appear that in the case of amorphous manganese there is less quenching of the electron orbits than in the other forms. The value obtained for the effective atomic moment, 3·09 Bohr magnetons, agrees well with the value 3·05 given by Sadron(13), who finds that the manganese atom dissolves in manganese-nickel and manganese-cobalt alloys with a mean moment of this amount.

The value of Δ, 1540°, is remarkably high, and is very interesting in view of the value, 1720°, recently calculated by Néel(10). He finds that a series of solid solutions of manganese in copper and silver obey Curie-Weiss laws, the appropriate values of the Curie constant in each case being directly proportional to the concentration of manganese. Hence the Curie constant for 1 gm.-mol. of manganese dissolved under these conditions can be computed. It is thus found that the manganese atom possesses an effective magnetic moment of 4·81 Bohr magnetons, which is considerably higher than any values hitherto recorded.

Néel(11) has recently given a theory of the constant paramagnetism of metals based on classical ideas. He supposes that in a specimen of pure manganese a negative molecular field causes alternate atoms to set with their moments oppositely directed so that the resultant magnetic moment of the specimen is zero. An applied magnetic field produces only a slight departure from this arrangement, and at high temperatures T greater than Δ the susceptibility is given by

$$\chi = C/T + \Delta;$$

while at low temperatures T less than Δ the susceptibility is given by

$$\chi_0 = C/3\Delta,$$

i.e. we have a constant susceptibility. Assuming that the value of χ_0 is $9{\cdot}9 \times 10^{-6}$ and taking his calculated value of C, Néel finds Δ to be 1720° from the last expression. It is therefore of some interest that a form of manganese exists which obeys a Curie-Weiss law even at low temperatures with Δ equal to 1540°.

The highest experimental values for the quantity Δ are found in the case of platinum. Föex and Miss Collet(2) have recently shown that platinum specimens behave in a very complicated way in different magnetic fields and over different ranges of temperature, and give different values for the effective magnetic moment for Δ. To a first approximation Δ was here directly proportional to the value of the Curie constant. The highest value of Δ, 3420°, was found with a specimen which possessed a moment of 13 Weiss magnetons over the range -180 to $-110°$ C. Consequently, beyond a suggestion that a high value of Δ means the presence of an internal molecular field which opposes magnetization, no satisfactory explanation of these high values can at present be given.

The modern quantum theory of the magnetic properties of metals has recently been discussed in an illuminating manner by Stoner(15). Paramagnetism, on this theory, is attributed to electron spins. In the absence of a magnetic field the number of spins pointing in one direction is balanced by an equal number pointing in the opposite direction. The establishment of a magnetic field causes an increase in the number pointing along the lines of force. Now according to the Fermi-Dirac statistics, this increase can only occur by the transition of electrons from cells of lower to cells of higher energy. In a metal the conduction electrons must be regarded as occupying a series of energy bands, and accordingly the possibility of transfer depends on the energy-separation between successive bands.

The comparatively high susceptibilities of the transition elements may be considered to be due to a narrowing of these energy bands, which in these atoms are derived from five *d* and one *s* states. Stoner suggests that these bands probably overlap, and perhaps should more properly be treated as one composite band in which the number of electrons able to contribute to the paramagnetism is equal to the number of unpaired electron spins in the atom. From these views is derived an expression for the susceptibility of manganese which requires that the variation with temperature over the range investigated in the above experiments should be of the order of 2 per cent, assuming the critical temperature, i.e. the temperature above which the conduction electrons may be taken to obey classical laws, to be 3150° K. Our experiments may therefore be interpreted as showing that the critical temperature must be much lower than this, and that the composite band must be very narrow.

§6. ACKNOWLEDGMENTS

We wish to thank Prof. E. N. da C. Andrade for the facilities afforded us. The cost of the magnet⁽¹⁾ used in this investigation was fully covered by a grant from the Government Grants Committee of the Royal Society, to whom we express our thanks.

REFERENCES

(1) L. F. Bates and B. J. Lloyd-Evans. *Proc. phys. Soc.* **45**, 425 (1933).
(2) G. Föex and Miss Collet. *J. Phys. Radium*, **2**, 290 (1931).
(3) H. Freese. *Phys. Z.* **29**, 191 (1928).
(4) Marie Gayler. *J. Iron Steel Inst.* **115**, 393 (1929).
(5) R. Hadfield, C. Chéveneau and C. Géneau. *Proc. roy. Soc.* A, **94**, 65 (1917).
(6) K. Honda. *Sci. Rep. Tokio Imp. Univ.* **1**, 1 (1912).
(7) H. Honda and T. Soné. *Sci. Rep. Tokio Imp. Univ.* **2**, 27 (1913).
(8) T. Ishiwara. *Sci. Rep. Tokio Imp. Univ.* **1**, 51 (1912).
(9) P. Kapitza. *Proc. roy. Soc.* A, **131**, 266 (1931).
(10) M. L. Néel. *J. Phys. Radium*, **3**, 160 (1932).
(11) M. L. Néel. Thèse, Strasbourg (1932).
(12) H. Owen. *Ann. Phys.*, Lpz., **37**, 657 (1912).
(13) C. Sadron. *C. R. Acad. Sci.*, Paris, **192**, 1311 (1931), and **193**, 1070 (1931).
(14) Y. Shimizu. *Sci. Rep. Tokio Imp. Univ.* **19**, 411 (1930).
(15) E. C. Stoner. *Magnetism and Matter*, Chap. xiv (1934).
(16) J. H. van Vleck. *Electric and Magnetic Susceptibilities*, p. 225 (1932).
(17) E. Vogt. *Ann. Phys.*, Lpz., **14**, 1 (1932).
(18) E. Vogt. *Sci. Progr. Twent. Cent.* **27**, 110 (1933).
(19) P. Weiss. *Congrès International d'Electricité*, Section 1, Rapport 15, p. 26 (1932).
(20) Mary A. Wheeler. *Phys. Rev.* **41**, 351 (1932).

612.843.3

THE COLOUR-VISION CHARACTERISTICS OF TWO TRICHROMATS

By W. D. WRIGHT, D.Sc., Ph.D., A.R.C.S., D.I.C.

AND

F. H. G. PITT, A.R.C.S., D.I.C., B.Sc.
(Imperial College of Science and Technology)

Received September 14, 1934. *Read in title December* 7, 1934.

ABSTRACT. The authors' luminosity curves, trichromatic coefficients, mixture curves, complementary wave-lengths and their mixture in terms of luminosity, their hue-discrimination curves and some data on saturation discrimination have been obtained and the results tabulated. The importance, for visual research, of complete data for individual observers, rather than mean curves for different groups of observers, is emphasized.

§ 1. INTRODUCTION

In the discussion on a recent paper on hue-discrimination⁽⁵⁾, the question was raised of the relation between the discrimination function and the colour-mixture data for individual eyes. This raises the larger question of the need, in visual research, of having complete information for each observer, rather than a number of mean curves, especially when these have been obtained for very different groups of observers. This need is of no particular concern in colorimetry, but visual problems can undoubtedly be attacked more profitably by investigating each individual observer completely, relating his various curves with each other, and then comparing the complete characteristics of different observers. This has led us to collect all the data we have for our own eyes and present it in a form suitable for comparing with other results as they become available. We believe this is the first time such complete information has been published for even one eye, but it is to be hoped that data for other observers will be forthcoming in the near future.

It will be realized that the larger part of the tables and curves given here do not represent new observations, but results that have been extracted from previous publications. This applies to the trichromatic coefficients, the hue-discrimination curves and, indirectly, the complementary colours. The luminosity curves and the saturation data represent new observations.

One or two points of interest are noted, but the results have not, as yet, been subjected to any detailed analytical study.

§ 2. APPARATUS

The apparatus used for all the measurements has been the Wright colorimeter, and as this has been described in detail elsewhere(3) no further description is given. The uses of the apparatus have also been given previously, so that, where further reference has been desirable in this paper, it has been made as brief as possible.

§ 3. LUMINOSITY

The evaluation of a luminosity curve involves two distinct measurements. We have first to compare the luminosities of narrow bands of a given spectrum and, secondly, we have to determine the relative amounts of energy in each band. There are three possible ways in which the first observations may be obtained, by direct comparison, by flicker, or by means of a trichromatic colour-match. The first method is useless for accurate work, the second is satisfactory if somewhat lengthy and tedious, while the third is very convenient and straightforward, provided the relative luminosities of the three primaries can be found. The latter method was adopted, as it was particularly convenient with our colorimeter, but it is surprising that the method is not more generally used in heterochromatic photometry. It should be very useful in these days, when gas discharge lamps of various colours are being used on such a large scale for illumination purposes.

The energy in the spectrum viewed in the colorimeter was computed from a knowledge of the colour temperature of the pointolite lamp, from measurements of the absorption of a filter that was used and whose transmission was not strictly uniform throughout the spectrum, from an approximate correction for the absorption in the prisms and lenses of the colorimeter, and from a further correction to allow for the varying width of spectrum transmitted through the exit pupil of the instrument.

The results for W.D.W. and F.H.G.P. are given in the second columns of tables 1 and 2 and in figures 1 and 6 respectively. They have been corrected to give the relative luminosity of an equal-energy spectrum and the ordinates have been adjusted so that the maximum luminosity is in each case 1·000. For W.D.W. the maximum is at 0·558 μ., while for F.H.G.P. it occurs at 0·562 μ.; these values compare with a maximum at 0·555 μ. for the standard luminosity curve. In the blue, both curves are somewhat higher than the normal curve, apparently owing to a less dense macular pigment, as will be shown below from the position of the white points in the colour triangle.

§ 4. TRICHROMATIC COEFFICIENTS

The trichromatic coefficients were obtained exactly as described elsewhere(4) and further description is unnecessary. It has, however, to be decided in what form the results shall be expressed. In 1931, the Commission Internationale de l'Éclairage defined the colorimetric characteristics of a standard observer in terms of three monochromatic radiations, 0·70 μ., 0·5461 μ. and 0·4358 μ., while the units of the

Wave-length	Relative luminosity of equal-energy spectrum	Trichromatic coefficients			Mixture curves for equal-energy spectrum			Complementary wave-lengths		Luminosity of λ to be mixed with unit luminosity of λ'	Hue-discrimination; just-noticeable wave-length differences	Saturation-discrimination; luminosity of white to be mixed with unit luminosity of λ	Wave-length
λ (μ.)		B (0·46 μ.)	G (0·53 μ.)	R (0·65 μ.)	L_B	L_G	L_R	λ (μ.)	λ' (μ.)		(μ.)		λ (μ.)
0·40	0·010	1·022	−0·054	0·036	0·028	−0·030	0·012	0·40	0·5693	0·0160	—	—	0·40
0·41	0·014	1·021	−0·053	0·034	0·036	−0·038	0·016	0·41	0·5696	0·0164	—	—	0·41
0·42	0·019	1·020	−0·051	0·033	0·047	−0·047	0·019	0·42	0·5698	0·0175	—	—	0·42
0·43	0·027	1·019	−0·048	0·031	0·061	−0·057	0·023	0·43	0·5700	0·0183	0·0076	—	0·43
0·44	0·037	1·018	−0·046	0·028	0·080	−0·071	0·028	0·44	0·5702	0·0196	0·0031	—	0·44
0·45	0·056	1·016	−0·034	0·018	0·098	−0·064	0·022	0·45	0·5704	0·0248	0·0040	—	0·45
0·46	0·085	1·000	0·000	0·000	0·085	0·000	0·000	0·46	0·5722	0·0437	0·00335	0·128	0·46
0·47	0·133	0·955	0·080	−0·035	0·062	0·094	−0·023	0·47	0·5758	0·100	0·00235	—	0·47
0·48	0·186	0·861	0·209	−0·070	0·039	0·187	−0·040	0·48	0·5822	0·245	0·00160	—	0·48
0·49	0·254	0·678	0·424	−0·102	0·023	0·281	−0·050	0·49	0·6030	0·795	0·00118	0·031	0·49
0·50	0·409	0·380	0·755	−0·135	0·012	0·452	−0·055	—	—	—	0·00118	—	0·50
0·51	0·604	0·170	0·958	−0·128	0·006	0·653	−0·055	—	—	—	0·00142	0·024	0·51
0·52	0·772	0·064	1·010	−0·074	0·002	0·808	−0·038	—	—	—	0·00190	—	0·52
0·53	0·874	0·000	1·000	0·000	0·000	0·874	0·000	—	—	—	0·00235	0·033	0·53
0·54	0·951	−0·018	0·934	0·084	−0·0009	0·900	0·052	—	—	—	0·00240	—	0·54
0·55	0·994	−0·020	0·851	0·169	−0·0010	0·882	0·113	—	—	—	0·00204	—	0·55
0·56	0·998	−0·021	0·761	0·260	−0·0012	0·818	0·181	—	—	—	0·00165	0·024	0·56
0·57	0·982	−0·018	0·662	0·356	−0·0010	0·730	0·253	0·4320	0·57	0·0185	0·00140	—	0·57
0·58	0·904	−0·015	0·543	0·472	−0·0008	0·579	0·326	0·4770	0·58	0·190	0·00115	0·030	0·58
0·59	0·781	−0·010	0·416	0·594	−0·0005	0·407	0·376	0·4860	0·59	0·41	0·00100	—	0·59
0·60	0·639	−0·008	0·302	0·706	−0·0003	0·250	0·389	0·4896	0·60	0·69	0·00095	—	0·60
0·61	0·518	−0·004	0·198	0·806	−0·0001	0·143	0·375	0·4910	0·61	1·00	0·00115	—	0·61
0·62	0·400	−0·003	0·123	0·880	−0·0001	0·071	0·329	0·4924	0·62	1·28	0·00155	0·024	0·62
0·63	0·284	−0·002	0·066	0·936	0·0000	0·028	0·256	0·4926	0·63	1·50	0·00210	—	0·63
0·64	0·181	−0·001	0·029	0·972	0·0000	0·008	0·173	0·4928	0·64	1·64	0·00285	—	0·64
0·65	0·111	0·000	0·000	1·000	0·0000	0·000	0·111	0·4930	0·65	1·72	0·0060	0·028	0·65
0·66	0·0686	0·000	−0·013	1·013	0·0000	−0·0014	0·070	—	—	—	—	—	0·66
0·67	0·0408	0·000	−0·018	1·018	0·0000	−0·0011	0·0419	—	—	—	—	—	0·67
0·68	0·0205	0·000	−0·022	1·022	0·0000	−0·0007	0·0212	—	—	—	—	0·036	0·68
0·69	0·0082	0·000	−0·026	1·026	0·0000	−0·0003	0·0085	—	—	—	—	—	0·69
0·70	0·0024	0·000	−0·029	1·029	0·0000	−0·0001	0·0025	0·4932	0·70	1·89	—	0·116	0·70
1931 C.I.E. S_B source		0·421	0·363	0·216	Relative luminosities of trichromatic units of primaries								
λ=0·4940 μ.		0·558	0·558	−0·116	0·46 μ.	0·53 μ.	0·65 μ.						
λ=0·5825 μ.		−0·014	0·507	0·507	0·051	1·000	0·648						

Table 2. Colour characteristics for F.H.G.P.

Wave-length	Relative luminosity of equal-energy spectrum	Trichromatic coefficients			Mixture curves for equal-energy spectrum			Complementary wave-lengths		Luminosity of λ to be mixed with unit luminosity of λ′	Hue-discrimination; just-noticeable wave-length differences	Saturation-discrimination; luminosity of white to be mixed with unit luminosity of λ	Wave-length
λ (μ.)		B (0·46 μ.)	G (0·53 μ.)	R (0·65 μ.)	L_B	L_G	L_R	λ (μ.)	λ′ (μ.)		(μ.)		λ (μ.)
0·40	0·012	1·028	−0·054	0·026	0·031	−0·028	0·009	0·40	0·5698	0·0175	—	—	0·40
0·41	0·016	1·027	−0·052	0·025	0·039	−0·036	0·013	0·41	0·5700	0·0180	—	—	0·41
0·42	0·020	1·027	−0·049	0·022	0·048	−0·041	0·013	0·42	0·5702	0·0189	—	—	0·42
0·43	0·026	1·026	−0·044	0·018	0·057	−0·044	0·013	0·43	0·5704	0·0204	0·0050	—	0·43
0·44	0·035	1·022	−0·037	0·015	0·065	−0·042	0·012	0·44	0·5707	0·0244	0·00275	—	0·44
0·45	0·045	1·016	−0·026	0·010	0·068	−0·031	0·008	0·45	0·5712	0·0306	0·00273	—	0·45
0·46	0·054	1·000	0·000	0·000	0·054	0·000	0·000	0·46	0·5721	0·0463	0·00260	0·033	0·46
0·47	0·073	0·938	0·082	−0·020	0·032	0·050	−0·009	0·47	0·5750	0·112	0·00189	—	0·47
0·48	0·120	0·829	0·221	−0·050	0·024	0·115	−0·019	0·48	0·5818	0·283	0·00146	—	0·48
0·49	0·196	0·647	0·445	−0·092	0·017	0·211	−0·032	0·49	0·6088	1·01	0·00118	0·035	0·49
0·50	0·336	0·387	0·729	−0·116	0·011	0·367	−0·042	—	—	—	0·00104	—	0·50
0·51	0·508	0·164	0·928	−0·092	0·0053	0·542	−0·039	—	—	—	0·00103	0·039	0·51
0·52	0·674	0·058	0·990	−0·048	0·0021	0·697	−0·025	—	—	—	0·00128	—	0·52
0·53	0·786	0·000	1·000	0·000	0·0000	0·786	0·000	—	—	—	0·00156	0·028	0·53
0·54	0·873	−0·023	0·952	0·071	−0·0011	0·829	0·045	—	—	—	0·00168	—	0·54
0·55	0·956	−0·024	0·866	0·158	−0·0013	0·845	0·112	—	—	—	0·00159	—	0·55
0·56	0·998	−0·019	0·771	0·248	−0·0012	0·810	0·189	—	—	—	0·00136	0·031	0·56
0·57	0·987	−0·015	0·668	0·347	−0·0009	0·717	0·271	0·410	0·57	0·0180	0·00115	—	0·57
0·58	0·908	−0·011	0·544	0·467	−0·0006	0·561	0·348	0·4780	0·58	0·234	0·00100	0·030	0·58
0·59	0·800	−0·009	0·418	0·591	−0·0005	0·396	0·405	0·4856	0·59	0·51	0·00085	—	0·59
0·60	0·665	−0·008	0·300	0·708	−0·0003	0·245	0·420	0·4886	0·60	0·80	0·00085	—	0·60
0·61	0·506	−0·006	0·202	0·804	−0·0002	0·130	0·376	0·4901	0·61	1·03	0·00089	—	0·61
0·62	0·371	−0·004	0·116	0·888	−0·0001	0·057	0·314	0·4915	0·62	1·18	0·00104	0·019	0·62
0·63	0·261	−0·003	0·060	0·943	−0·0001	0·021	0·240	0·4920	0·63	1·31	0·00131	—	0·63
0·64	0·177	−0·001	0·021	0·980	0·0000	0·005	0·172	0·4922	0·64	1·42	0·00195	—	0·64
0·65	0·110	0·000	0·000	1·000	0·0000	0·000	0·110	0·4924	0·65	1·51	0·00275	0·015	0·65
0·66	0·062	0·000	−0·010	1·010	0·0000	−0·0009	0·063	—	—	—	0·0050	—	0·66
0·67	0·0312	0·000	−0·018	1·018	0·0000	−0·0008	0·032	—	—	—	—	—	0·67
0·68	0·0159	0·000	−0·021	1·021	0·0000	−0·0005	0·0164	—	—	—	—	0·026	0·68
0·69	0·0073	0·000	−0·022	1·022	0·0000	−0·0002	0·0075	—	—	—	—	—	0·69
0·70	0·0029	0·000	−0·023	1·023	0·0000	−0·0001	0·0030	0·4926	0·70	1·60	—	0·026	0·70
1931 C.I.E. S_B source		0·422	0·367	0·211	Relative luminosities of trichromatic units of primaries								
λ = 0·4940 μ.		0·552	0·552	−0·104	0·46 μ.	0·53 μ.	0·65 μ.						

three stimuli were taken as equal when mixed together so as to match the colour possessed by a hypothetical source having an equal-energy spectral radiation.

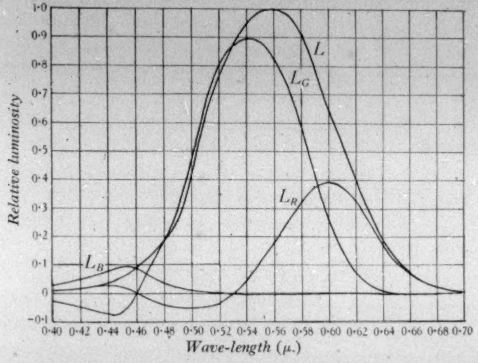


Figure 1. Luminosity curve and mixture curves for equal-energy spectrum (W. D. W.).

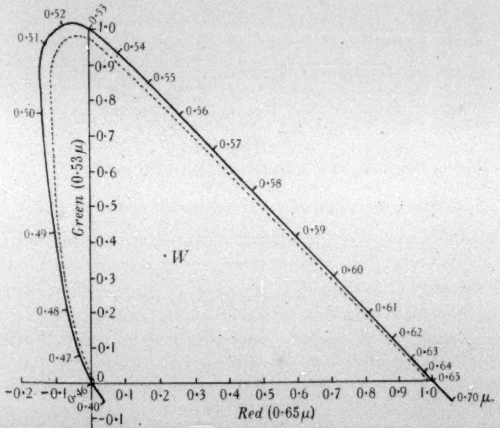


Figure 2. Spectral locus and white point (W. D. W.), plotted in colour triangle on W. D. W. system. Dotted curve shows the locus of points having a just-noticeable difference in saturation from the spectral colours.

Should the coefficients given here be expressed in the same framework? It would be convenient, for instance, to be able to compare the authors' coefficients with those of the standard observer and it is certainly desirable to keep the number of different reference systems down to a minimum. On the other hand, the standard observer has been defined for use in colorimetry and is only indirectly of interest to the physiological side of the subject. The system of units used by Wright$^{(4)}$ in 1929 has

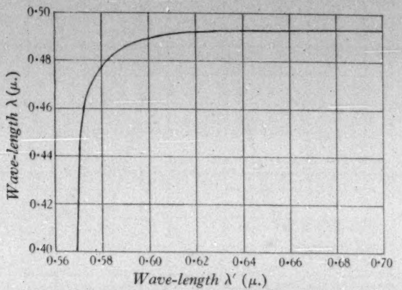


Figure 3. Curve showing relation between complementary wave-lengths λ and λ′ (W. D. W.).

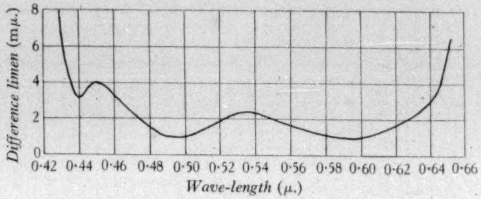


Figure 4. Hue-discrimination curve (W. D. W.).

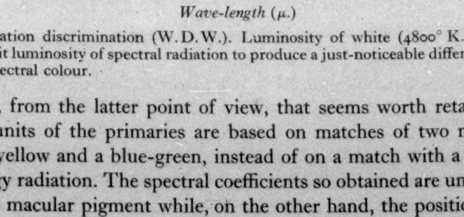


Figure 5. Saturation discrimination (W. D. W.). Luminosity of white (4800° K.) which has to be added to unit luminosity of spectral radiation to produce a just-noticeable difference in saturation from the spectral colour.

an advantage, from the latter point of view, that seems worth retaining. In this method the units of the primaries are based on matches of two monochromatic radiations, a yellow and a blue-green, instead of on a match with a source of continuous-energy radiation. The spectral coefficients so obtained are unaffected by the density of the macular pigment while, on the other hand, the position of the white point in the colour triangle is a direct indication of the yellowness of the pigment. This feature, we feel, should be retained, and although it could be done while still retaining the primaries prescribed by the Commission Internationale de l'Éclairage, there would be little point in doing so if the units of the primaries were defined on another system. In fact, a hybrid arrangement of this sort would probably cause more confusion than two distinct systems. We therefore decided to retain the monochromatic basis, which we may perhaps be allowed to distinguish as the W. D. W. system, using the primaries we have employed for the past seven years, namely 0·65 μ., 0·53 μ., and 0·46 μ. In the W. D. W. system, the red and green units are adjusted to be equal when matched against a yellow, 0·5825 μ., and the blue and green units are corrected so that equal amounts of 0·46 μ. and 0·53 μ. are required in the match on 0·4940 μ.

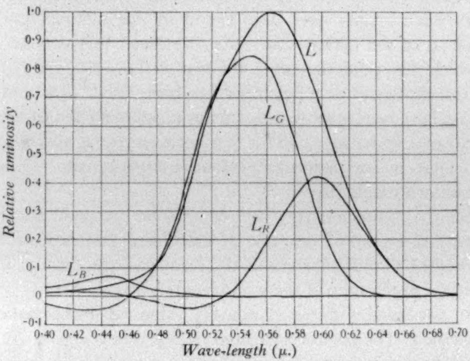


Figure 6. Luminosity curve and mixture curves for equal-energy spectrum (F. H. G. P.).

The authors' coefficients are given in columns 3, 4 and 5 of tables 1 and 2 and are shown in the colour triangle in figures 2 and 7. It is interesting to note that in the blue-green region, the slightly larger negative red coefficients for W. D. W. correspond to larger positive red values in the violet part of the spectrum. Also, the two white points, which by pure coincidence happen to be practically identical, are both on the blue side of the corresponding point for the standard observer, whose coefficients on this system are $R = 0{\cdot}249$, $G = 0{\cdot}399$, $B = 0{\cdot}352$. This indicates that the authors' macular pigment is rather less dense than that of the standard observer, and this would in turn account for the higher ordinates found at the blue end of the luminosity curve. The illuminant used to provide this white point was the S_B source, having a colour temperature of 4800° K., of the Commission Internationale de l'Éclairage (1931).

§ 5. MIXTURE CURVES

The mixture curves, which merely express the colour-mixture data in terms of luminosity, have been computed from the luminosity curve, the trichromatic coefficients and the relative luminosities of the three primaries, in the usual manner. The data is given for an equal-energy spectrum in columns 6, 7 and 8 of the two tables and diagrammatically in figures 1 and 6. The three mixture curves when added together give, of course, the luminosity curve. Some small differences in the two sets of results, such as the shape of the red curves, the position of the green maxima and, most of all, the differences at the blue end, are of interest.

§ 6. COMPLEMENTARY COLOURS

Once the spectral locus and the white point in the colour triangle are known, it is possible, from the geometry of the triangle, to locate complementary wave-lengths. If a straight line passing through the white point intersects the spectral locus at wave-lengths λ and λ', then these two wave-lengths are complementary. Pairs of complementary wave-lengths have been found in this manner and are given in columns 9 and 10 of the tables. The values obtained will depend on the white radiation to which they refer and, as in the rest of the work, the S_B source of the Commission Internationale de l'Éclairage has been used. For convenience, the complementaries have been given for radiations at intervals of 0·010μ., both for the range between 0·40μ. and 0·49μ. and for that between 0·57μ. and 0·65μ. The green region has, of course, no complementary colour in the spectrum. In figures 3 and 8, the results are shown graphically.

Not only can the complementary wave-lengths be deduced, but the proportion in which they must be mixed can also be calculated and expressed in terms of luminosity, provided the luminosities of the primaries are known. The procedure involves merely the application of the well-known centre-of-gravity property of the colour triangle, combined with the luminosity values of trichromatic units of the radiations concerned. The results calculated in this way are given in column 11 of the tables and show the intensity of λ required to be mixed with unit intensity of λ' to match the white illuminant.

§ 7. HUE-DISCRIMINATION

The determination of the authors' hue-discrimination curves has been described recently⁽⁵⁾ and the results given here have been extracted directly from the previous paper. In column 12 of the tables the wave-length difference required to produce a just-noticeable colour-difference has been tabulated for intervals of 0·010μ. through the spectrum. The results are shown graphically in figures 4 and 9.

§ 8. SATURATION-DISCRIMINATION

We should have liked to include some measurements on the relative saturation of the spectral colours, similar to the curves given by Priest and Brickwedde⁽²⁾ and by Martin, Warburton and Morgan⁽¹⁾. These investigators have obtained curves showing the luminosity required by each monochromatic radiation to produce, when mixed with a given luminosity of white, a just-noticeable colour-difference from white. In figure 11 a mean curve taken from the paper by Martin, Warburton and

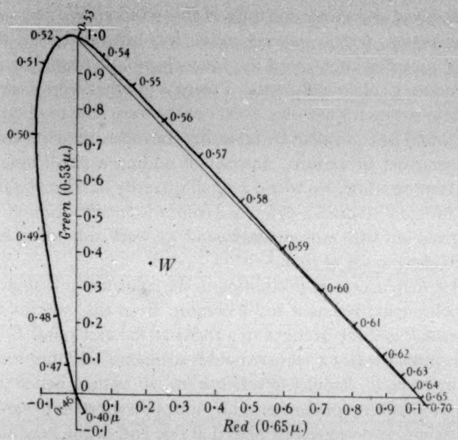


Figure 7. Spectral locus and white point (F. H. G. P.), plotted in colour triangle on W. D. W. system.

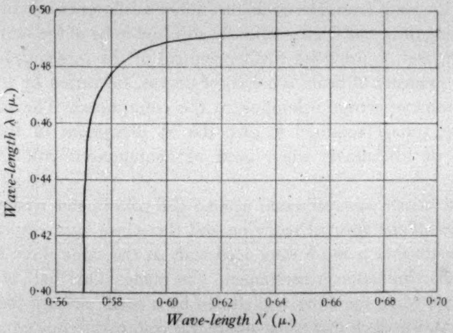


Figure 8. Curve showing relation between complementary wave-lengths λ and λ′ (F. H. G. P.).

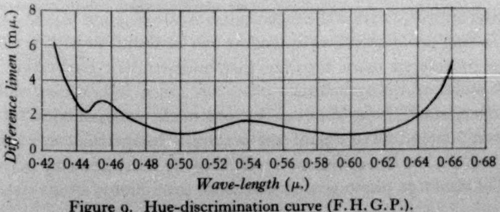


Figure 9. Hue-discrimination curve (F. H. G. P.).

Morgan is reproduced and shows one form of this relation. We can see, for instance, that the yellow region, $0{\cdot}57\mu$., is very much less saturated than the rest of the spectrum, since a greater intensity of the monochromatic beam has to be added to the white to produce a colour-difference. Pressure of other work prevented us from determining these curves for our own eyes, but we were able to obtain some results that we hoped would be equivalent by reversing the measurements and thus simplifying the experimental procedure. Instead of adding a small amount of monochromatic radiation to white, we added a small quantity of white light to the monochromatic radiation, to produce a difference from the spectral colour. This could be fairly easily carried out with our apparatus and we were able to get some results in a comparatively short space of time.

The optical system necessary in addition to the colorimeter is shown in figure 12. The two monochromatic beams *a* and *b* coming from the colorimeter pass into a photometer prism *P* and are brought to a focus at the exit pupil *E* by the lens *O*. When the eye is placed at *E*, a 2°-square field is seen, one half illuminated by beam and the other by beam *b*. Parallel to this system, an optical bench *B* was fixed on which a 1000-watt lamp *L* could be moved by the observer so as to vary the illumination on a magnesium oxide screen *M* placed at one end of the bench. A glass plate *G* mounted at 45° to the direction of beam *b* enabled a fraction of the light from the magnesium screen to be mixed with beam *b*, and the observer merely had to adjust the position of the lamp *L* until he could just notice a difference in saturation between *a* and *b*, assuming these had been previously adjusted to be of the same wave-length. Two sectors S_1 and S_2 provided further control of the intensities of the field if required. The intensity of beam *a* could, of course, be varied by the observer by using the photometer wedge belonging to the colorimeter. The 1000-watt lamp was run at the voltage required to give the S_B illuminant of the Commission Internationale de l'Éclairage, when used in conjunction with the appropriate filter at *F*.

The optical bench was calibrated against the colorimeter wedges and the relative intensities of the spectral radiation and the white light could be calculated readily. To ensure that *a* and *b* were accurately at the same wave-lengths, a preliminary hue-discrimination measurement was made. One half of the field was illuminated by λ, while the other was altered by a small amount to a wave-length $(\lambda+\delta\lambda_1)$, until $\delta\lambda_1$ was such that a just-noticeable colour-difference between the two halves of the field could be seen. The measurement was repeated on the short-wave side of λ, to a wave-length $(\lambda-\delta\lambda_2)$, and the value was again recorded. To a first approximation $\delta\lambda_1=\delta\lambda_2$, so that the mean value of $(\lambda+\delta\lambda_1)$ and $(\lambda-\delta\lambda_2)$ will be λ and the correct setting for identical wave-lengths can be found to within 1 or 2 A. This method was considered more accurate than independent calibration of the two wave-length scales on the apparatus.

The intensity at which the observations were made was maintained at an approximately constant level by the use of the sectors. Except for the wave-lengths of $0{\cdot}70\mu$. and $0{\cdot}46\mu$., the intensities never differed by more than 25 per cent from a mean value of about 50 photons.

The steps as measured for 10 wave-lengths are given in column 13 of the tables. The values give l_w, the small white luminosity that must be added to unit intensity of each spectral radiation to produce a just-noticeable saturation difference. The results are shown graphically in figures 5 and 10. l_w

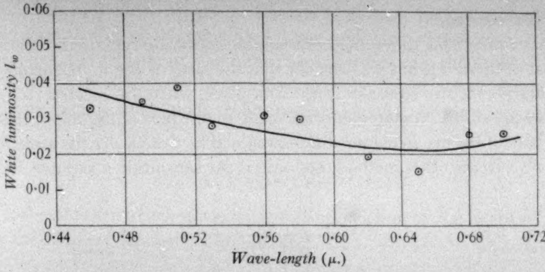


Figure 10. Saturation discrimination (F.H.G.P.). Luminosity of white (4800° K.) which has to be added to unit luminosity of spectral radiation to produce a just-noticeable difference in saturation from the spectral colour.

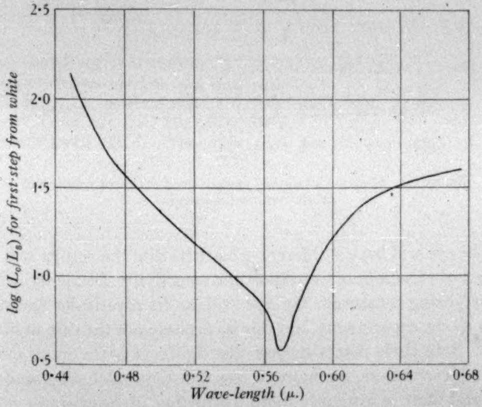
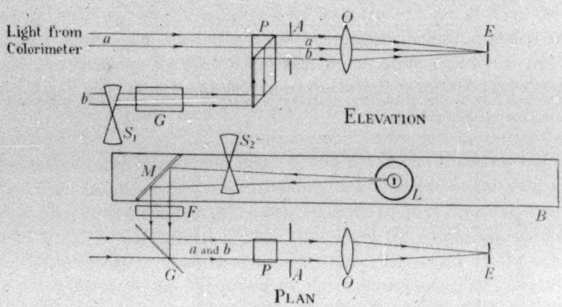


Figure 11. Saturation discrimination—mean curve taken from Martin, Warburton and Morgan, showing the relation between spectral radiation (luminosity L_s) which, when mixed with white (luminosity L_w), gives a colour (luminosity L_c, $L_c = L_s + L_w$) which has a just-noticeable saturation difference from white.

The rather unexpected result has been obtained that l_w remains very nearly constant for all wave-lengths. This result was quite definitely confirmed for W.D.W. on two or three occasions, and although time did not permit the observations for F.H.G.P. to be repeated so thoroughly, there was every reason to believe that more extensive observations would have produced equally constant results. The two end values for W.D.W. were higher than the rest, possibly but not necessarily as a result of the lower intensity-level. It was of some interest to find that the intensity seemed to play a bigger part in changing the size of saturation-step than had been found in the case with hue-discrimination. This may be associated with the decreasing saturation (increasing whiteness) known to occur as the intensity of a radiation is increased.

There is a surprising difference between the saturation functions of figures 5 and 10 and that of figure 11, but we consider that the latter is the more correct measure of the saturation for two reasons. In the first place, when the spectral radiations are added to white, all the measurements are made in one part of the colour triangle and the sensitivity of the eye to colour-differences will be practically the same for all the observations. In the other method, not only is the saturation a variable, but at each wave-length the eye will have a different sensitivity and the results will measure the combined effect of variations of saturation and sensitivity. Secondly, when the steps from white are being measured, the eye will be in practically the same state of adaptation for all the experiments; but this is certainly not the case in the method we have adopted. Both these points suggest that figure 11 is the more reliable saturation function. Nevertheless, the more or less constant value we found for l_w has to be explained and there is evidently room for further investigation.



ELEVATION

PLAN

Figure 12. Diagram of optical system used for measurement of saturation discrimination.

The saturation-steps plotted in the colour triangle are of some interest and we have therefore given, in figure 2, the first step from the spectral locus for W.D.W.'s colour triangle. It must be remembered when using this diagram that the colour difference is only a liminal one when measured along a line passing through the white point. It would be incorrect, for instance, to draw a normal to the spectral locus and regard the intersections with the locus and the dotted line as giving two points having a just-noticeable colour-difference.

§ 9. CONCLUSIONS

Several points of interest have been brought out in the course of the paper and these are probably sufficient to demonstrate the advantage of a complete series of results for each individual observer. The value of each item is undoubtedly enhanced by a knowledge of the remaining items, and when comparison is made with other observers, the manner in which the characteristics vary should be of particular value. These variations and their connexion with the more marked abnormalities of colour-vision, such as anomalous trichromatism, should also be of special interest.

§ 10. ACKNOWLEDGMENT

The authors desire to acknowledge their appreciation of the financial assistance given by the Medical Research Council.

REFERENCES

(1) L. C. Martin, F. L. Warburton and W. J. Morgan. *Spec. Rep. Ser. med. Res. Coun.* No. 188.
(2) I. G. Priest and F. G. Brickwedde. *J. opt. Soc. Amer.* **13**, 306 (1926).
(3) W. D. Wright. *Trans. opt. Soc.* **29**, 225 (1927–8).
(4) W. D. Wright. "A re-determination of the trichromatic coefficients of the spectral colours." *Trans. opt. Soc.* **30**, 141 (1928–9).
(5) W. D. Wright and F. H. G. Pitt. *Proc. phys. Soc.* **46**, 459 (1934).

551.594.12

THE POTENTIAL ACQUIRED IN THE NATURAL ELECTRIC FIELD BY A VERTICAL ROD STANDING ON THE GROUND, INSULATED AT THE BOTTOM AND CARRYING A COLLECTOR AT THE TOP

By L. H. G. DINES, M.A.

Received October 18, 1934. *Read in title December* 7, 1934.

ABSTRACT. This paper investigates the potential of a small cylindrical rod standing upon, and insulated from, a charged plane, and provided with a small collector at the top. Data of a direct observational comparison between such a rod and a standard piece of apparatus are given. The main part of the paper consists of a theoretical determination of the potential of the rod by means of an approximate solution of the integral equation involved. The investigation is extended to cover a certain range of values of the ratio of the diameter to the height of the rod. An attempt is made to indicate the distribution of charge on the surface of the rod in a special case. An estimate is made of the subsidiary error involved when the collector takes the form of a horizontal fuze which shortens as it burns away.

---

§ 1. STATEMENT OF THE PROBLEM

The measurement of the electric stress near the surface of the ground is a problem which has exercised the minds of meteorologists for several decades. The physical quantity which it is desired to measure is the vertical gradient of electric potential over a large horizontal surface, but in the earlier attempts the measuring apparatus employed itself distorted the field so much that the results obtained were subject to large error.

Several improved forms of apparatus were devised about twenty-five years ago, and one which was adopted by the late G. W. Walker was installed by him at Eskdalemuir Observatory, Dumfriesshire, and has been in use there since. It consists of a thin insulated vertical rod, of which the lower end passes through a small hole in a large horizontal plane into an underground chamber, in which sits an observer with an electrometer connected to the rod. A collector is attached to the top of the rod in the form of a small burning fuze placed horizontally.

It was originally assumed that when equilibrium had been attained the potential of the rod would be very nearly the same as the undisturbed potential at points in the horizontal plane passing through the top of the rod, and it is the object of the present paper to test the truth of this assumption.

Under normal conditions the surface of the ground is negatively charged, and a qualitative consideration of the problem shows that as the rod is mainly at a higher potential than its surroundings it must have positive charges on its surface, and its potential must therefore be higher than the undisturbed potential in the neighbourhood of the burning fuze. In a paper by G. M. B. Dobson(1), published about twenty years ago, reference is made in a footnote to a comparison between the vertical rod and another form of apparatus employing a horizontal rod, which showed that the former gave readings about 4 per cent higher than the latter. Unfortunately very few details of this test are available.

Figure 1 shows in diagrammatic form the shape of the equipotential surfaces in the neighbourhood of a plain insulated rod fitted with an ideal small collector at the extreme top. It is demonstrated later that the critical equipotential which passes through the top of the rod takes a form in that region which departs little from that of a cone. This fact suggested that if the long thin fuze commonly used as a collector were placed on the slope so as to form a generator of the cone, then the presence of the fuze would not disturb the configuration of the equipotential surfaces, and observations so made would give results identical with those made with a plain rod and an ideal small collector. On this basis, a series of comparisons were made at Kew Observatory between a vertical rod with an inclined fuze and a modern form of apparatus substantially free from systematic error. This work was done subsequently to the theoretical investigation, but it seems best to give an account of it first.

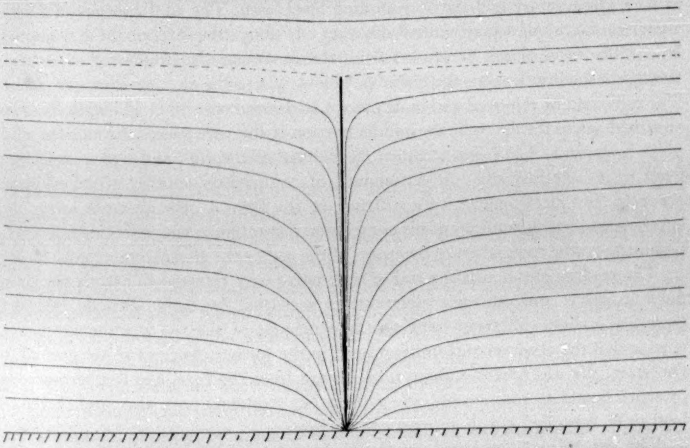


Figure 1. Approximate diagrammatic section of the equipotential surfaces in the neighbourhood of an insulated vertical rod 100 cm. in height and 4 mm. in diameter, standing upon a charged horizontal surface and provided with a collector at the top.

§ 2. AN EXPERIMENTAL DETERMINATION

Early in the year 1933 a series of simultaneous readings of potential were made, use being made on the one hand of a vertical rod fitted with an inclined fuze, and on the other a long insulated horizontal wire provided with a fuze in the middle. On theoretical grounds it is demonstrable that the latter form of apparatus is not subject to appreciable error due to distortion of the electric field. A description of it is given by Dr R. E. Watson(3).

The site of the comparison was the roof of the underground electrical laboratory at Kew Observatory and the surrounding level lawn. The roof consists of a horizontal surface 10 m. square raised about 25 cm. above the level of the surrounding lawn. The vertical rod protruded from the roof near the middle of the square, through a small hole from the chamber below, where the electrometer was placed. The rod could be removed and in its place a horizontal wire 18 m. in length could be stretched across the flat roof; the middle portion of the wire passed through the same point in space as had been occupied by the top of the rod, and had a small fuze fixed to it concentrically. As a standard of comparison another stretched wire, observer, and electrometer were situated on the lawn a little distance away; the latter observer took readings of his electrometer simultaneously with every observation made on the roof, whether by means of the rod or the alternative stretched wire.

The reading of potential by rod or alternative wire referred to exactly the same point in space, and the same electrometer was used for both. Strictly speaking there was a slight difference between the exposures of the rod and the alternative wire in that the observer was underground in the former case and above ground in the latter. He was however about 9 m. distant from the fuze, and furthermore his presence would to some extent affect the other stretched wire also, and therefore tend to be cancelled out in a direct comparison.

The vertical rod employed was approximately 1 m. in height and 4·1 mm. in diameter. The theoretical investigation which follows shows that for such a rod the fuze should be inclined at 7° above the horizontal, and it was accordingly so fixed. At the point where it joined the vertical rod its axis was 1 m. above the roof, its diameter was 4½ mm. and initial length from 7 to 10 cm., burning away to nothing at the end of the test in each case.

Seven complete comparative tests were made. In each case simultaneous readings were made by both observers at intervals of half a minute. Table 1 sets out the mean results of each portion of each comparison, the nomenclature adopted being as follows:

V_r V_r is the potential of the vertical rod;
V_a V_a the potential of the stretched wire over the roof; and
V_c V_c the potential of the control wire over the lawn.

The values of V_r/V_a have been derived from the mean values of V_r/V_c and V_a/V_c in the rows above. The whole series was made under good conditions in dry weather

Table 1

Date, 1933	April 4	April 7	April 7	April 8	April 10	April 11	April 11	Mean
No. of readings	16	15	15	20	20	23	15	
Mean of V_r/V_e	1·08	1·09	1·11	1·10	1·08	1·09	1·10	1·09
No. of readings	18	11	17	17	21	23	19	
Mean of V_a/V_e	1·02	1·05	1·02	1·06	1·06	1·02	1·03	1·03
V_r/V_a	1·06	1·04	1·09	1·04	1·06	1·06	1·08	1·06

with a gentle breeze. The scatter of the individual values in the last row is too great to allow us to say with precision what the mean value of a very large number of similar observations would be, but it is fairly safe to say that it must lie between 1·05 and 1·07.

§ 3. THE THEORETICAL INVESTIGATION

In dealing with the potential of the rod from the theoretical point of view it is necessary to simplify the conditions. The fuze will be supposed to be replaced by a particle of ionizing material attached to the tip of the rod, and the shape of the latter will be regarded as a cylinder of length h and diameter $2a$, having plane h, a orthogonal ends. The cylinder stands vertically upon, but is insulated from, an infinite horizontal plane which is uniformly charged with a surface density q. The q only physical property of the collector which we need consider is that it produces a state of equilibrium in which the surface density of charge at the top of the rod is zero. The writer does not know of any published solution of this problem, nor of the more tractable case when the rod takes the form of a thin prolate spheroid. It can however be attacked by methods of approximation, and the method which has been adopted here was suggested by the description of a somewhat similar process given by L. F. Richardson in a paper entitled "How to solve differential equations approximately by arithmetic(2)."

The problem may be stated analytically thus. Take the axis of the rod as axis of Y, the positive direction being downwards with the origin at the top of the rod. Let σ be the surface density of the charge on the rod at the distance y from the top. σ, y At the top end of the rod the surface density is zero: at the bottom end, which is nearly in contact with the infinite plane, there will be a large charge. The latter will have in close proximity a corresponding charge on the plane, nearly equal and opposite to it, and the external effect of such pairs falls off as a high power of the distance. Its effect will be ignored in the present investigation. We are therefore solely concerned with charges on the cylindrical surface. The surface density of the charge on the infinite plane is disturbed by the presence of the charged rod; its effect may be dealt with by the method of images, so that we shall envisage the problem as that of a charged rod above the plane, the image of the rod below, and the plane having a uniform surface density q.

y' Consider any point on the axis of Y inside the rod, and let y' denote its distance from the origin. Since it is inside a conductor the electric stress there is zero, and therefore from a consideration of all the elemental charges in the system and the vertical stresses which they individually produce at the point considered, we have the following integral equation which is true for all values of y' between zero and h.

$$2\pi a\left\{\int_0^h \frac{\sigma(y'-y)}{((y'-y)^2+a^2)^{\frac{3}{2}}}\,dy+\int_0^h \frac{\sigma(2h-y'-y)}{((2h-y'-y)^2+a^2)^{\frac{3}{2}}}\,dy\right\}=4\pi q.$$

We have to determine σ as a function of y, and thence determine the potential of the rod relative to the plane. The process being numerical it is necessary to start with a rod of definite dimensions. Assume a rod 1 m. in length and 4 mm. in diameter; it is also convenient to assume that q is equal to $-\frac{1}{4}\pi$, because this gives a vertical potential-gradient equal to unity in absolute units above the plane, and positive upwards. This case having been worked out any other will follow proportionally. The rod is supposed to be divided into n equal portions, and the assumption is made that $d\sigma/dy$ is constant throughout each portion. Our hypothetical distribution therefore involves discontinuities in $d\sigma/dy$ but not in σ. We shall assume that this hypothetical distribution makes the electric stress zero at certain convenient points on the axis of the rod, and then work out numerically for several values of n the values of the distribution constants which are required to meet these conditions. Lastly the potential of the rod will be evaluated, and from a series of values obtained from increasing values of n an approximation to the limiting value when n is infinite will be made. It will be obvious that these hypothetical distributions could not exist in equilibrium for finite values of n; they are of value only as a means of approach to the limit, and it is shown later that it is not necessary to carry the process very far before the limiting value can be estimated.

The convenience of the numerical process is somewhat dependent on choosing the best points on the axis at which to equate the electric stress to zero, and it was found that to obtain approximation to a limit it was necessary to work with the middle points of the sections of the rod.

V In the case of any of our hypothetical distributions let V be the potential relative to the infinite plane at any point in space due to all the charges in the system, and let V_r be the potential at the same point due to the charges on the rod and its image alone. Then, for any of our chosen points on the axis of the rod,

V_r

$$\partial V/\partial y=0 \quad \text{and} \quad \partial V_r/\partial y=+1.$$

$\sigma_a, \sigma_b, \sigma_c, \sigma_d$ At the upper ends of each of the n sections into which the rod is divided, beginning from the top, we may denote the values of σ thus: Zero, σ_a, σ_b, σ_c, σ_d, σ_e σ_e, etc. Starting from this we may then express the value of $\partial V_r/\partial y$ for the middle point of each section, and by equating each of them to $+1$ obtain thereby n equations from which to determine σ_a, σ_b, σ_c, σ_d, σ_e, etc. Lastly the potential of the top of the rod can be computed from the known values of σ_a, σ_b, σ_c, σ_d, σ_e etc. The whole process is very easy so long as n is small, but becomes tedious when n exceeds three.

In evaluating $\partial V_r/\partial y$ at any of the proposed points it is more convenient to regard it as a reversed electric stress than as a function of potential. By the aid of a few formulae, which are worked out below, each section of the rod and image can be dealt with in succession, and it is evident that each complete expression must have $2n$ elements.

The formulae required. Suppose a cylindrical shell of external diameter $2a$ having its axis as axis of Y. The shell is supposed to be charged to a surface density σ, where σ obeys the relation $\sigma = s + k.y$, where s and k are constants.

s, k

Consider the electric force, P_0, at O the origin of co-ordinates, due to a finite portion of the shell lying between the limits defined by $y = y_1$ and $y = y_2$, where y_2 is algebraically greater than y_1. By symmetry P_0 acts along OY, and its value in the positive direction of y is given by the integral

P_0

y_2, y_1

$$P_0 = -2\pi a \int_{y_1}^{y_2} \sigma y \,(a^2+y^2)^{-\frac{3}{2}} dy.$$

Since $\sigma = s + ky$ this becomes

$$P_0 = -2\pi a s \int_{y_1}^{y_2} y \,(a^2+y^2)^{-\frac{3}{2}} dy - 2\pi a k \int_{y_1}^{y_2} y^2 \,(a^2+y^2)^{-\frac{3}{2}} dy.$$

Writing $\tan\theta = a/y$ and integrating, we have

$$P_0 = 2\pi s \int_{\theta_1}^{\theta_2} \cos\theta \, d\theta + 2\pi a k \int_{\theta_1}^{\theta_2} \cos\theta \cot\theta \, d\theta$$

$$= 2\pi s (\sin\theta_2 - \sin\theta_1) + 2\pi a k (\cos\theta_2 - \cos\theta_1 + \log(\tan\tfrac{1}{2}\theta_2 \cot\tfrac{1}{2}\theta_1)).$$

If σ_1 and σ_2 be the values of σ corresponding with y_1 and y_2 respectively, we shal have

$$\sigma_1 = s + ky_1 \quad \text{and} \quad \sigma_2 = s + ky_2,$$

whence

$$s = \frac{\sigma_1 y_2 - \sigma_2 y_1}{y_2 - y_1} \quad \text{and} \quad k = \frac{\sigma_2 - \sigma_1}{y_2 - y_1}.$$

Substituting in the above we have

$$P_0 = \frac{2\pi}{y_2 - y_1} \{(\sigma_1 y_2 - \sigma_2 y_1)(\sin\theta_2 - \sin\theta_1) + a(\sigma_2 - \sigma_1)(\cos\theta_2 - \cos\theta_1 + \log(\tan\tfrac{1}{2}\theta_2 \cot\tfrac{1}{2}\theta_1))\}.$$

If a/y be small so that we may neglect squares and higher powers, and if y_1 and y_2 have the same sign, this reduces to

$$P_0 = \frac{2\pi a}{y_2 - y_1} \left\{ \left(\frac{\sigma_2}{y_2} - \frac{\sigma_1}{y_1} \right)(y_2 - y_1) + (\sigma_1 - \sigma_2) \log\frac{y_2}{y_1} \right\} \qquad \text{......(1).}$$

In the special case when $y_1 = 0$ we have

$$P_0 = \frac{2\pi}{y_2} \{\sigma_1 y_2 (\sin\theta_2 - 1) + a(\sigma_2 - \sigma_1)(\cos\theta_2 + \log\tan\tfrac{1}{2}\theta_2)\}.$$

Again, if a/y_2 be small and y_2 be positive, this reduces to

$$P_0 = \frac{2\pi a}{y_2}\left\{\sigma_1\left(1-\frac{y_2}{a}\right)+(\sigma_2-\sigma_1)\left(1-\log\frac{2y_2}{a}\right)\right\} \qquad \ldots\ldots(2).$$

If the portion of the shell considered lie on the negative side of the origin, equation (2) must be modified; in that case y_2 is zero and y_1 is negative.

We then have

$$P_0 = \frac{2\pi a}{y_1}\left\{\sigma_2\left(1+\frac{y_1}{a}\right)-(\sigma_2-\sigma_1)\left(1-\log\frac{-2y_1}{a}\right)\right\} \qquad \ldots\ldots(3).$$

If the origin be at the centre of a portion of the shell whose total length is $2y_1$, then by a similar process we have, when a/y_1 is small,

$$P_0 = \frac{2\pi a}{y_1}(\sigma_1-\sigma_2)\left(\log\frac{2y_1}{a}-1\right) \qquad \ldots\ldots(4).$$

ϕ_0 Again, the potential ϕ_0 at the origin O, due to the portion of the shell lying between the limits y_1 and y_2, is given by the integral

$$\phi_0 = 2\pi a\int_{y_1}^{y_2}\sigma\,(y^2+a^2)^{-\frac{1}{2}}\,dy.$$

Whence, proceeding as before, we obtain the equation

$$\phi_0 = \frac{2\pi a}{y_2-y_1}\left[(\sigma_2-\sigma_1)\left\{(y_2^2+a^2)^{\frac{1}{2}}-(y_1^2+a^2)^{\frac{1}{2}}\right\}-(\sigma_1 y_2-\sigma_2 y_1)\log\left(\tan\tfrac{1}{2}\theta_2\cot\tfrac{1}{2}\theta_1\right)\right].$$

If a/y be small and y_1 and y_2 have the same sign, this reduces to

$$\phi_0 = 2\pi a\left\{\frac{\sigma_1 y_2-\sigma_2 y_1}{y_2-y_1}\log\frac{y_2}{y_1}+(\sigma_2-\sigma_1)\right\} \qquad \ldots\ldots(5).$$

Again, in the special case when a/y_2 is small, y_1 is zero and y_2 is positive, we have

$$\phi_0 = 2\pi a\left\{\sigma_1\log\frac{2y_2}{a}+(\sigma_2-\sigma_1)\left(1-\frac{a}{y_2}\right)\right\} \qquad \ldots\ldots(6).$$

To find the potential due to the shell at an external point T, take axes as before and suppose that the co-ordinates of T are given by $x=x$ and $y=0$. Suppose also that x is large compared with a, so that we may take the surface charge on the shell to be concentrated along the axis, that is we may take a/y to be negligible from the start.

The potential at T due to that part of the rod lying between the limits $y=y_1$ and $y=y_2$ is then given by the expression

ϕ_T

$$\phi_T = 2\pi a\int_{y_1}^{y_2}\sigma\,(y^2+x^2)^{-\frac{1}{2}}\,dy.$$

Integrating, we obtain

$$\phi_T = \frac{2\pi a}{y_2-y_1}\left[(\sigma_1 y_2-\sigma_2 y_1)\log\left(\tan\tfrac{1}{2}\theta_1\cot\tfrac{1}{2}\theta_2\right)+(\sigma_2-\sigma_1)\left\{(y_2^2+x^2)^{\frac{1}{2}}-(y_1^2+x^2)^{\frac{1}{2}}\right\}\right] \qquad \ldots\ldots(7),$$

where $\theta_1 = \tan^{-1} x/y_1$ and $\theta_2 = \tan^{-1} x/y_2$.

§ 4. NUMERICAL EXAMPLES

By the aid of the general equations (1) to (7) the numerical equations may be written down in any specific case. As has previously been stated, attention has been confined to the case of a rod 1 m. in height and 4 mm. in diameter. The calculations have been made for values of n from 1 to 5 inclusive. In each case the constants σ_a, σ_b, σ_c, σ_d, etc. have been computed, and thence the potential at the top end of the rod determined. Lack of space prevents the setting out of the work in detail, but a summary of the results is given in table 2. Each row in the table deals with one definite hypothetical case. The last column gives the potential computed for the top end of the rod; this is a somewhat arbitrary proceeding because if the computation had been made for some other point on the rod the figures would not necessarily have been quite the same, though they must approximate to the same limit as n is increased. There are advantages in computing for the top, because that is a region of zero stress at the surface of the rod.

Table 2. Computed values of the density of the surface charge on a rod 1 m. high and 4 mm. in diameter in the five hypothetical cases considered, with the resultant potential at the top of the rod relative to the plane

Distance from top of rod (cm.)	0	20	25	33·33	40	50	60	66·67	75	80	100	Potential of the top of the rod (V)
	(Surface density of charge on rod) × 4π											
$n=1$	0	—	—	—	—	—	—	—	—	—	104·90	106·417
$n=2$	0	—	—	—	—	47·70	—	—	—	—	123·10	106·090
$n=3$	0	—	—	32·32	—	—	—	66·25	—	—	136·69	106·134
$n=4$	0	—	24·71	—	—	48·95	—	—	77·29	—	148·13	106·184
$n=5$	0	20·13	—	—	39·25	—	59·77	—	—	85·10	158·32	106·224

We see from table 2 that the potential of the rod appears to be somewhat in excess of 106. If the rod were removed the potential at the level of the top would be 100, so that the disturbance produced by the rod is of the order of 6 per cent or more. To determine the limiting value we may plot the potential against $1/n$ and extrapolate the curve. The figures in the last column of table 2 are shown plotted as small circles in figure 2. Such a method is not very satisfactory, and it is better to proceed as follows. There is every reason to suppose that the potential converges to a limit as n is increased indefinitely; call this limit V' and we may assume that the difference between V' and the estimate obtained by using any particular value of n will take the form

V'

$$\alpha/n+\beta/n^2+\gamma/n^3+\delta/n^4+\dots,$$

$\alpha, \beta, \gamma, \delta$ where α, β, γ, δ, etc. are independent of n. From the data of table 2 five equations may now be formed from which to determine V' and the first four coefficients α, β, γ and δ, provided that it may be assumed that the rest of the series is insignificant.

Thus,

$$\begin{aligned}
106{\cdot}417 - V' &= \alpha + \beta + \gamma + \delta,\\
106{\cdot}090 - V' &= \tfrac{1}{2}\alpha + \tfrac{1}{4}\beta + \tfrac{1}{8}\gamma + \tfrac{1}{16}\delta,\\
106{\cdot}134 - V' &= \tfrac{1}{3}\alpha + \tfrac{1}{9}\beta + \tfrac{1}{27}\gamma + \tfrac{1}{81}\delta,\\
106{\cdot}184 - V' &= \tfrac{1}{4}\alpha + \tfrac{1}{16}\beta + \tfrac{1}{64}\gamma + \tfrac{1}{256}\delta,\\
106{\cdot}224 - V' &= \tfrac{1}{5}\alpha + \tfrac{1}{25}\beta + \tfrac{1}{125}\gamma + \tfrac{1}{625}\delta.
\end{aligned}$$

Solving, we find $\alpha = -1{\cdot}653$, $\beta = +2{\cdot}287$, $\gamma = -1{\cdot}358$, $\delta = +0{\cdot}669$ and $V' = 106{\cdot}47$.

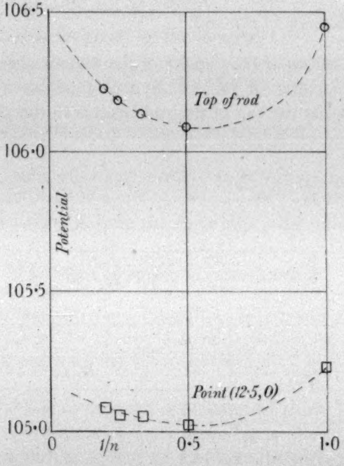


Fig. 2. Estimates of the potential at the top of the rod, and at a point level with the top but 12·5 cm. distant from it, based on various values of $1/n$.

As a check the same process was followed using only the first three terms and the first four equations; in this case the value of V' was found to be 106·45.

As the validity of the limiting value is vital to the investigation it has been examined in another manner. If the hypothetical distributions of charge given in table 2 agree closely with the real distribution, then it must follow that the values of the function $\partial V_r/\partial y$ derived from them must approximate to unity at all points on the axis of the rod. The values when $n = 5$ are shown plotted in figure 3, from which it will be seen that the agreement is reasonably good except at the extreme top and near the bottom. The same procedure was carried out for smaller values of

n, and it was found that both discrepancies decrease as n increases. Discrepancies at the bottom have little influence in any case, while at the top was found to decrease in a manner roughly proportional to $1/n$. Hence there is no reason to doubt that the calculated potential will converge steadily to a limit as n increases, and also that the most important terms in the expression for the difference between the estimate and its limit will be those involving the smaller powers of n.

When the figures obtained above have been rounded off, the best estimate of V' which the present investigation can give is 106·5 for a rod 100 cm. in height and 4 mm. in diameter.

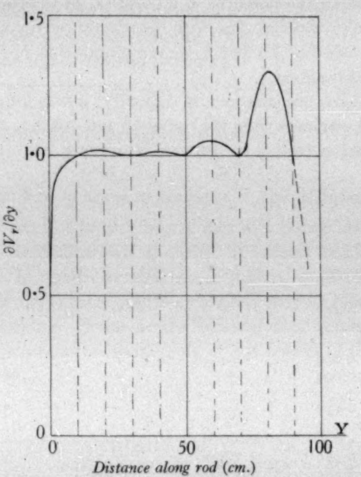


Figure 3. Values of $\partial V_r/\partial y$ along the axis of the rod, computed for the case in which $n=5$.

§ 5. AGREEMENT WITH EXPERIMENT

The value of the potential just obtained may be compared with the experimental result on page 221. The agreement is not perfect, but is good enough to check the general accuracy of both methods. The theoretical investigation suggests that the experimental value must be too low; there is no reason why it should not be so, since it is always difficult to avoid appreciable casual error in observations of potential gradient. The scatter among the seven experimental determinations on page 221 is such that the true mean may be as great as 106·5 for all we can tell.

There is a difference between the practical case and the theoretical one in that in the former the rod passes through a hole in the plane, whereas in the latter it does not. If the hole be small this is immaterial, as the difference in the electric stress in the two cases is entirely negligible except in the immediate neighbourhood of the hole, where it does not affect the calculations.

§ 6. THE SHAPE OF THE EQUIPOTENTIAL SURFACE PASSING THROUGH THE TOP OF THE ROD

It appears from the foregoing calculations that a reasonably good estimate of the potential of the rod can be made even when n is as small as 2 or 3. We may therefore with confidence make use of the hypothetical distributions to determine the shape of the equipotential surfaces near the top of the rod. Since there is no charge at the top of the rod there can be no question of discontinuities in that region. Taking axes as before, consider the potential at a point (x, y) in space near the top of the rod; using equation (7) we may deal with the several portions of the rod and image as before and work out V_r. To find V we remember that $V = V_r + (100 - y)$, where y is measured in centimetres.

The work is tedious but presents no difficulty; it was found convenient to take points at horizontal distances from the rod of $\frac{1}{8}h$ and $\frac{1}{16}h$, i.e. 12·5 and 6·25 cm. The potentials are set out in table 3 for various values of n.

Table 3. Values of V_r and V at points in space at and near the top of the rod in various cases of hypothetical distribution of charge on the rod.

Co-ordinates (x, y) (cm.)	(0, 0)	(6·25, 0)	(6·25, −1·25)	(12·5, 0)	(12·5, −2·5)
$n=1$ V_r	6·417	—	—	5·233	4·880
V	106·417	—	—	105·233	107·380
$n=2$ V_r	6·090	—	—	5·023	4·670
V	106·090	—	—	105·023	107·170
$n=3$ V_r	6·133	—	—	5·055	—
V	106·133	—	—	105·055	—
$n=4$ V_r	6·185	5·602	5·370	5·057	4·685
V	106·185	105·602	106·620	105·057	107·185
$n=5$ V_r	6·225	—	—	5·085	—
V	106·225	—	—	105·085	—

Taking the case in which $n=4$ we can find the points at the same potential as the top of the rod by interpolation on the ordinates x = 6·25 and x = 12·5. They are found to be (6·25, −0·71) and (12·5, −1·32). These are nearly collinear with the top of the rod. Since the distribution in this case has been shown to be not far from the true one, it may be inferred that the curvature of the equipotential through the top of the rod in the actual case will be very nearly the same as in that just found.

It is next necessary to determine the limiting value of V_r at the point (12·5, 0) when n is infinite. The five estimates in table 3 are plotted as small squares in figure 2. The means of calculation employed (slide-rule and four-figure tables) were not quite accurate enough to give a smooth curve, but as the principles involved are exactly the same as before we may draw a smooth curve through the points, as far as they allow, similar in shape to the curve above. This curve may then be extrapolated to the ordinate through the origin by exactly the same process as before, save that as we have already smoothed the curve we can only profitably work from four points instead of five. Solving the necessary equations, the limiting value of V is found to be 105·17.

Lastly we require the limiting value of the vertical potential gradient near the point (12·5, 0). Table 3 shows that there is little difference between the computed values whether n be 1, 2 or 4, and therefore we may take the value in the latter case as indicating the actual value sufficiently closely for the purpose, namely a gradient of 2·13 in 2·5 cm.

From the data of the last three paragraphs the equipotential through the top of the rod in the actual case may be determined thus:

The potential at the top of the rod is 106·47.

On the ordinate $x=12\cdot5$ this potential occurs at the point given by the equation $y=-(1\cdot30/2\cdot13)\times2\cdot50$ cm. That is, $y=-1\cdot53$ cm.

Allowing for curvature the intermediate point is

$$x=6\cdot25 \text{ cm.},\quad y=-0\cdot87 \text{ cm.}$$

Since the equipotential surface is symmetrical with regard to the axis of Y, its form near the top of the rod is very nearly that of a cone, having a semi-vertical angle of 83° from the vertical drawn upwards. It follows that if a thin conductor not more than 13 cm. in length be attached to the rod at this angle at the top, it will lie in an equipotential surface, and the collector may be placed at any point on it without altering anything. Hence a fuze fulfilling these conditions will maintain the rod at the same potential independently of the length of the fuze as it burns away.

§ 7. THE ACTUAL DISTRIBUTION OF THE CHARGE ON THE SURFACE OF THE ROD

It remains to discuss the actual distribution of charge on the rod. Five approximations have been worked out for a rod 100 cm. by 4 mm., and in figure 4 three of them are shown plotted corresponding with values of n of 1, 3 and 5; the other two are very similar but have been omitted for the sake of clearness. The question is, what form would the curve take if n were very great?

Since the various approximations have all yielded estimates of the potential which have been shown to be not far from the limiting value, they cannot be very far from the true distribution, except at the bottom end which has not been explored. Taking the case when $n=5$ as the best, certain refinements suggest themselves which may be applied to it. In the first place, an examination of figure 3 shows that at the extreme top of the rod the gradient of charge must be somewhat steeper than has been postulated. Also at 20 cm. above the plane rather too steep a gradient has been obtained. Furthermore, table 2 shows that at 50 cm. above the plane the density is on the whole tending to increase as n increases from 2 to 5, while the density at the bottom if plotted against $1/n$ suggests an asymptotic curve when *n* is large. These several points have been incorporated in the upper curve in figure 4. The noteworthy feature is the close approximation of the curve to a straight line over the upper two-thirds of the length of the rod.

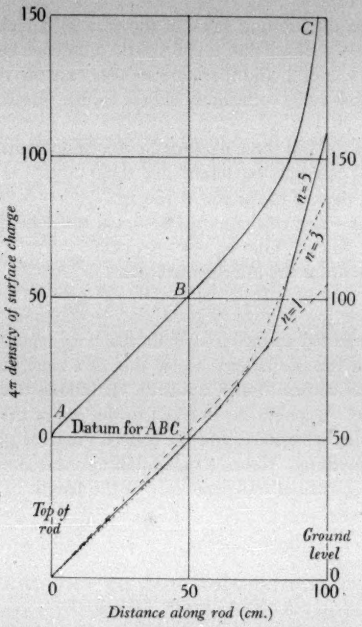


Figure 4. Graph of the product 4π into the density of the surface charge on a rod 100 cm. in height and 4 mm. in diameter, for various values of *n*. The limiting case is shown above as *ABC*.

§ 8. EFFECT OF THE DIMENSIONS OF THE ROD

The results so far obtained have been based on a rod of definite dimensions; we have next to consider how far they are generally applicable. It will be seen that in the equations for the determination of P_0 on pages 223, 224 each term is homogeneous and of the first power in σ, but of zero dimensions in the linear factors. Also in the equations for the determination of ϕ_0 the terms are of the first power in both σ and the linear factors. The same conditions apply to the electric stresses and potentials due to the charged plane. It follows that provided the ratio of the diameter to the height of the rod remains the same, identical values of σ_a, σ_b, σ_c, σ_d, etc. will be obtained whatever the height of the rod may be. Equipotential surfaces will remain similar figures as the scale varies, but the absolute magnitude of the potential will be proportional to the linear dimensions. Hence for any rod for which the ratio of diameter to height is 0·004, all the preceding results are immediately applicable by the application of a simple dimensional factor to all the potentials.

If the ratio of diameter to height be varied, further examination is required. By means of the same methods as before the potentials were worked out for rods 100 cm. in height and 2 and 6 mm. in diameter. They are tabulated in table 4, in which for convenience the results previously obtained for a 4 mm. rod are also included. The limiting values of V_r (0), the value of V_2 at the top of the rod, have been determined from the four cases when $n=1$, 2, 3 and 4, respectively. The limiting value of V_r (0) for a 4-mm. rod is not exactly the same as that previously found by taking five values of n, but we may with little risk of error assume that though the actual values in the fifth row of table 4 are not the best available, yet the proportions between them are very nearly correct. Accordingly in the last row are entered the estimated values of the potential, based on the previously accepted value of 6·47 for the 4-mm. rod.

V_r (0)

Table 4. Potential V_r (0) for rods of various diameters

Diameter of rod	0·2 cm.	0·4 cm.	0·6 cm.
$n=1$	5·612	6·417	7·000
$n=2$	5·372	6·090	6·609
$n=3$	5·395	6·134	6·658
$n=4$	5·436	6·184	6·721
Limiting value of V_r (0) deduced from above	5·704	6·450	7·077
Estimated value of potential	5·72	6·47	7·10

The shape of the equipotential surface through the top of the rod can be worked out in detail as before, but it is not necessary to do so. The potential V_r at any point external to the rod is determined by means of equations similar to (7) on page 224, from which it is seen that V_r is proportional to $2a\sigma$, where σ is used in the general sense. If then in any special case all the σ's are altered in a certain ratio, the external potentials V_r will also alter in the same proportion.

In table 5 the figures are given for the surface distribution on the three rods in the hypothetical case when $n=4$. σ_{25} denotes the surface density at 25 cm. from the top, and so on.

Table 5. Surface density when $n=4$ on rods 100 cm. in height and of various diameters

Diameter of rod	0·2 cm.	0·4 cm.	0·6 cm.
σ_{25}	43·5	24·7	17·9
σ_{50}	86·3	49·0	35·4
σ_{75}	135·7	77·3	56·1
σ_{100}	246·4	148·1	111·9
$2a\,\sigma_{25}$	8·70	9·88	10·74

It will be seen that the distributions in the three cases are very nearly similar; the only appreciable discrepancy is at the bottom of the rod, where the effect on the potential is comparatively small; we may assume that much the same conditions exist in the actual case. The values of $2a\sigma_{25}$ in the bottom row may therefore be taken to represent the proportional magnitudes of the similar distributions of linear charge on the three rods. It follows that the external potentials V_r are proportional to the values of $2a\sigma_{25}$ in the bottom row. Now it will be seen from table 4 that when $n=4$, V_r (0) is also very nearly proportional to $2a\sigma_{25}$, so that the configurations of the isopleths of V_r passing through the tops of the rods are very nearly identical in the three cases. This result has been proved for the case of $n=4$, it can readily be worked out and shown to be true when $n=2$, and we may assume that it is true in the actual case also.

If now we repeat for rods 0·2 and 0·6 cm. in diameter the analysis made on page 228 we find the equipotential surfaces through the tops of the rods to be cones of semivertical angles 83·8° and 81·9° respectively; if fuzes be used with such rods the angle of inclination must be adjusted accordingly.

The results so far obtained may be put in the following general form. A vertical rod of height h and diameter $2a$ is exposed in the manner before described on an infinite charged horizontal plane, having a surface density of q. At the top is fixed a small collector, or a fuze whose length does not exceed $\frac{1}{8}h$, set at an angle of ψ ψ above the horizontal. The potential V' of the rod and the appropriate value of ψ are set out in table 6 for various values of $2a/h$.

Table 6

$2a/h$	0·002	0·004	0·006
$V'/4\pi qh$	− 1·0572	− 1·0647	− 1·0710
ψ	6·2°	7·0°	8·1°

§ 9. THE INFLUENCE OF THE FUZE AS COMMONLY EMPLOYED

When the vertical rod is used as a means of measuring the electric stress it is usual to fix the fuze horizontally, and in such a case the conditions are not quite the same as those which have been discussed above. With a rod 1 m. in height and 4 mm. in diameter the fuze is generally about 12·5 cm. in length and about 5 mm. in diameter, and burns away to nothing in the course of the observation. At the start the active end is situated at the point whose co-ordinates are (12·5, 0), where it will be seen from the results given on page 228 that the potential, apart from the presence of the fuze, is less than that at the tip of the rod by 1·30. It is not possible however to say that the equilibrium potential of the rod and fuze will therefore be less by this amount, because the field of force is altered by the presence of the fuze.

An approximation can be made in the following way. Roughly speaking the fuze lies in a potential-gradient along its axis of 0·104 per cm.; using equation (4) and the same methods as before, and assuming a uniformly increasing surface density of charge on the fuze from the burning tip to the junction with the top of the rod, we find that a fuze 4 mm. in diameter would have a surface density of $-0{\cdot}53/\pi$ at its junction with the rod, and one of 6 mm. a surface density of $-0{\cdot}40/\pi$. Considering the case of the 4-mm. fuze, since there cannot be a discontinuity as between fuze and rod we may assume that the charges on the rod have been modified in such a way that the surface density at the top is now $-0{\cdot}53/\pi$ instead of zero. As a rough approximation it may be assumed that this modification takes the form of a uniform reduction by $0{\cdot}53/\pi$ of the surface density all over the surface of the rod, which if anything will be an overestimate. Acting on this assumption and using the formulae and methods already given we may calculate by how much the potential of the rod and fuze is less than that of the original rod alone. The result is that for either a 4- or 6-mm. fuze the potential is about 0·65 less than the original potential at the point (12·5, 0).

On this basis the potential of the rod and fuze when the fuze is 12·5 in length will be $(106{\cdot}47-1{\cdot}30-0{\cdot}65)$, or 104·52, increasing to 106·47 as it just burns away. Taking the mean over the life of the fuze we have 105·5 as the mean potential of the rod and fuze. This result is not greatly dependent on the diameter of the fuze, and may be applied to any between 3 and 6 mm.

§ 10. CONCLUSION

The writer feels a certain amount of diffidence in presenting this approximate solution of a problem hitherto unsolved. Such solutions are liable to leave loose ends here and there, and details are inevitably open to criticism. The method of procedure however seems to have been justified by the agreement between the calculated potential and that found by direct experiment.

An interesting point arose in the course of the work in the choice of the best positions on the axis at which to equate $\partial V_r/\partial y$ with the electric intensity due to the undisturbed plane; the upper ends of each section were tried instead of the middle points, but the result failed to give an approximation to the limit with any reasonable value of n. The reason became apparent when a diagram similar to that of figure 3 was plotted; the curve showed great oscillations, and these became more pronounced as n increased. A few graphs drawn for several values of n on this principle provide an excellent criterion of the validity of the method of approximation employed.

A noteworthy feature of the method which was finally adopted is the comparative closeness of all the several estimates of the potential of the rod to the final limiting value. Even the worst of them, when $n=2$, departs from it by an amount which is less than the error which nearly always exists in any individual observational measurement of the potential-gradient. The simple case when $n=1$, the calculations of which only take a few minutes to perform, yields a surprisingly close estimate. Unfortunately this could not be predicted beforehand, but the fact may be useful in other applications of the same process.

§ 11. ACKNOWLEDGMENTS

The writer's thanks are due to Dr L. F. Richardson, F.R.S., and to Dr F. J. W. Whipple, both of whom made valuable suggestions in the course of the work.

REFERENCES

(1) GORDON DOBSON. *Geophys. Mem.* **1**, 161 (H.M. Stationery Office, 1914).
(2) L. F. RICHARDSON. *Math. Gaz.* p. 415 (July 1925).
(3) R. E. WATSON. *Geophys. Mem.* **5** (H.M. Stationery Office, 1929).

535·338.4:546.481.61

THE BAND SYSTEMS OF CADMIUM FLUORIDE

By R. K. ASUNDI, R. SAMUEL and M. ZAKI-UDDIN,
Muslim University, Aligarh

Received September 21, 1934. *Read in title December* 7, 1934.

ABSTRACT. Two emission band systems attributed to CdF are described and analysed. One, in the orange region, is probably due to a $^2\Pi \rightarrow {}^2\Sigma$ transition, and the other, in the yellow-green, to $^2\Sigma \rightarrow {}^2\Sigma$. The vibrational analyses suggest that the two systems have a common lower state, $^2\Sigma$, which is presumably the ground state. The origins of the two systems are roughly 16558 and 18871 cm.$^{-1}$ respectively, and the vibrational coefficients ω_e and $x_e\omega_e$ are of the orders 694·3 and 4·96 for the lower $^2\Sigma$ state, 704·4 and 5·74 for the $^2\Pi$ state, and 672·4 and 5·14 for the excited $^2\Sigma$ state. The heats and products of dissociation and the electronic structures of the three states of CdF are discussed.

§ 1. INTRODUCTION

LITTLE seems to be known about the spectra of cadmium fluoride. The spectrum of CdH is well known and the existence of certain bands supposed to be due to the molecules Cd_2, CdO, $CdCl_2$, $CdBr_2$ and CdI_2 as studied in emission in flame and discharge tube is also quoted in literature$^{(1)}$. Spectroscopically the CdF molecule should behave similarly to BeF, MgF, CaF, SrF and other similar molecules. The spectra of these molecules are studied and detailed vibrational analyses are available$^{(1)}$. Data on some of them have also been utilized recently by Lessheim and Samuel$^{(2)}$ in explaining the linkage between the valence electrons of the two constituent atoms. In each of these molecules the metal atom has two valence electrons, both in the *s* state, while the halogen has five *p* electrons in the outermost ring. Following Herzberg$^{(3)}$, these authors show that for the formation of a molecule in such cases the excitation of one of the *s* electrons of the metal atom to a *p* state is necessary before linkage can take place with one of the *p* electrons of the halogen. From the spectra of CaF and SrF etc. they show further that the anomalous terms of the alkaline earth atoms are involved in such of the higher electronic energy levels of these molecules as have a dissociation energy greater than in the ground level. It was thought that CdF may behave in a rather analogous way and an investigation of its spectrum was therefore undertaken.

§ 2. EXPERIMENTAL PROCEDURE

Pure cadmium fluoride was packed in bored copper rods, and also in Hilger's pure graphite rods in some cases, and the arc on 110 volts d.c. was struck between them. It was found that a current of about 5 amperes developed the bands with optimum strength. The bands were found to lie in the range $\lambda\lambda$ 6000–6300 and λλ 5300–5530; the entire quartz and glass regions up to 9000 A. were explored with low-dispersion instruments, but no other bands were developed. In their general appearance the bands are very similar to those of CaF.

The region to be explored having been ascertained, the bands were photographed with instruments of higher dispersion. A Zeiss 3-prism glass spectrograph which gives a dispersion of about 35 A./mm. between λλ 6600 and 6000 and about 25 A./mm. between λλ 5500 and 5300 was employed. A small concave grating with a radius of 124 cm. and 30,000 lines to the inch, having a dispersion of about 10 A./mm. in the first order, was also employed.

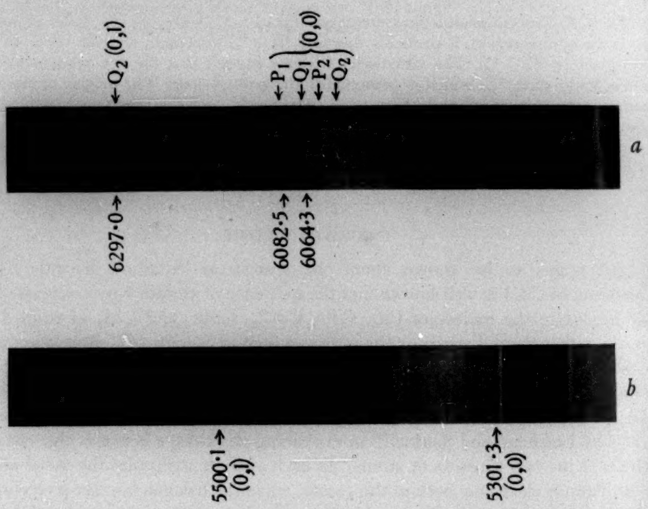


Figure 1.

For the final photographs three plates were taken for the orange and yellow-green bands with the 3-prism spectrograph and two with the grating for the orange. They were measured on a Zeiss-Abbé comparator readable directly to 0·001 mm. and estimable to 0·0001 mm.; in every case the cross-wire was focused on the most intense part of the band-head. The copper lines, the cadmium lines and some iron lines were used as standard wave-lengths. In the case of the prism spectrograms wave-lengths were calculated by means of the usual Hartmann dispersion formula. The mean of three measurements of wave-lengths calculated on three separate plates in the case of the prismatic spectrograms and two of the grating was taken. In most cases the agreement in wave-lengths measured on the different plates was satisfactory, the maximum divergence being 1 A. On account of the rather low dispersion used and the rather diffuse nature of some of the band-heads, very high accuracy cannot be claimed, the probable error being about 0·3 A. for the prism plates and 0·1 A. for the grating plates. However, in the entire region in which the bands lie, an error of 0·3 A. introduces an error of not more than one unit in wave-number values.

Figure 1 contains the reproductions on an enlarged scale of the various spectrograms taken. A glance at these pictures reveals the close similarity in the physical structure between these bands and those due to CaF; the bands in the orange, figure 1 *a*, are evidently similar to the red bands of CaF, and those in the yellow-green, figure 1 *b*, to the green bands of CaF. Analysis further shows some more similarity to which we shall come later. Tables 1 and 2 give the heads of the two different systems of the CdF bands arranged in the usual v' v'' notation.

Table 1. CdF yellow-green bands

v', v''	R_2 heads		R_1 heads		v', v''	R_2 heads		R_1 heads	
	λ_{air}	$\nu_{\text{vac.}}$	λ_{air}	$\nu_{\text{vac.}}$		λ_{air}	$\nu_{\text{vac.}}$	λ_{air}	$\nu_{\text{vac.}}$
0, 0	2 5301·3	18858	2 5299·3	18865	0, 1	1 5500·2	18176	1 5497·7	18184
1, 1	2 5307·0	18838	2 5305·5	18843	1, 2	1 5504·2	18163	0 5501·1	18173
2, 2	1 5313·2	18816	1 5311·3	18823	2, 3	1 5508·5	18149	1 5506·0	18157
3, 3	1 5319·5	18794	1 5317·8	18800	3, 4	1 5511·7	18138	0 5511·0	18141
4, 4	1 5325·2	18773	1 5323·7	18779	4, 5	1 5516·2	18123	1 6515·7	18125
5, 5	3 5334·3	18741	1 5329·7	18758	5, 6	1 5520·5	18109	0 5518·7	18115
6, 6	3 5340·6	18719	1 5336·2	18735	6, 7	0 5524·8	18095	0 5523·6	18099
7, 7	1 5346·9	18697	5 5342·9	18711	7, 8	0 5529·6	18079	1 5528·1	18084
8, 8	0 5353·8	18673	4 5349·2	18689	8, 9			5532·0	18071
9, 9	1 5360·4	18650	2 5355·6	18667					
10, 10	1 5367·0	18627							

Table 2. CdF orange bands

v', v''	Q_2 heads		P_2 heads		Q_1 heads		P_1 heads	
	λ_{air}	$\nu_{\text{vac.}}$	λ_{air}	$\nu_{\text{vac.}}$	λ_{air}	$\nu_{\text{vac.}}$	λ_{air}	$\nu_{\text{vac.}}$
0, 0	4 6036·9	16561	1 6053·3	16515	4 6064·3	16485	5 6082·5	16436
1, 1	3 6034·6	16567	5 6050·6	16523	4 6061·8	16492	2 6079·4	16444
2, 2	3 6032·5	16572	5 6048·6	16528	5 6060·0	16497	1 6077·7	16449
3, 3	2 6031·0	16576	2 6046·7	16533	3 6058·8	16500	1 6076·1	16453
4, 4	1 6029·3	16581	1 6045·0	16538	5 6057·6	16504	1 6074·6	16457
5, 5	1?6027·9	16585	0 6043·1	16543	3 6056·0	16508	1 6073·5	16460
6, 6					3 6054·7	16512	1 6072·4	16463
0, 1	5 6297·0	15876						
1, 2	(Masked)							
2, 3	4 6285·5	15905						
3, 4	5 6280·9	15917						
4, 5	5 6276·1	15929						
5, 6	4 6272·5	15938						
6, 7	4 6268·3	15949						

§ 3. DESCRIPTION OF THE BANDS

As is well known the wave numbers of the origins of bands belonging to the same electronic band system are expressed by the formula

$$\nu_0 = \nu_e + [\omega_e' (v' + \tfrac{1}{2}) - x_e' \omega_e' (v' + \tfrac{1}{2})^2] - [\omega_e'' (v'' + \tfrac{1}{2}) - x_e'' \omega_e'' (v'' + \tfrac{1}{2})^2].$$

If, as in most cases, only the heads of the bands are known, the same equation is satisfactory, provided the wave-number difference between the origin and the heads is constant for all the bands. In many cases however the situation is not very simple, and, according to the nature of the emitter, many characteristic behaviours of the system are met with. In this particular case, as in the case of CaF, SrF, CaCl, SrCl and other similar molecules, we find that only two sequences are developed, namely the 0, 0 and the 0, 1. This points clearly to a case in which the vibrational frequency and the anharmonic constant do not differ greatly in the various electronic levels. Further, the moments of inertia being also about the same, the peculiar physical structure of the bands pointed out by Johnson(4) obtains here. The similarity between these bands and those of the already known molecules of the alkaline earth halides has been of considerable help in arriving at a proper analysis of these bands.

In the yellow-green system we find that the 0, 0 sequence is very clearly developed. Each band-head is accompanied by a nearly equally intense band-head, as in the case CaF. This band system evidently belongs to the transition $^2\Sigma \rightarrow {}^2\Sigma$ in the molecule. The double-headed structure of the bands may be due, as Harvey(5) has pointed out in the case of CaF, to the two R branches (arising from a wide ρ-type doubling) each forming a head of its own. If, on the other hand, we regard one of these as a R head and the other as a Q head, as Johnson(4) does, it is rather difficult to explain how such strong Q branches can be present in a $\Sigma \rightarrow \Sigma$ transition. Quite empirically, therefore, we call the two heads R_1 and R_2. Along the sequence the band-heads go on diverging, and this is a clear indication that we are dealing with an excited level for which, though the frequency of vibration is smaller than the frequency of vibration of the unexcited state, the anharmonic constant is bigger; in this respect also these bands are perfectly analogous to the similar CaF band. The 0, 1 sequence is rather diffuse; but it has been possible to measure R_1 and R_2 heads. Photometric plates of the bands of this sequence were also taken on a Zeiss recording micro-photometer, which helped to identify the position of the bands with more exactness. Figure 2 is a reproduction of the photometric plate taken of this sequence. For the purpose of calculation the values measured on the photometric plates are utilized.

The orange system develops the 0, 0 sequence strongly and its structure clearly indicates that a $\Pi \longleftrightarrow \Sigma$ transition is involved. There are two Q heads and two P heads in the 0, 0 sequence. The 0, 1 sequence however develops only Q heads; this is a handicap for the analysis of the bands, but it seems fairly certain that this is the Q series corresponding to the Q_2 series of the 0, 0 sequence. The other arrangement of putting them with the Q_1 series changes the vibrational frequency by 75 cm.$^{-1}$, which is highly improbable because we should then have no similarity between the final level of this band system and that of the yellow-green system. The analysis was attempted with a view to correlate, if possible, these two electronic levels, and this we think is perfectly natural because of our knowledge of the behaviour of similar molecules. The appearance of only one Q branch in this sequence is rather surprising but by no means exceptional. We know, for example, that in the analogous bands of CaF(4) the 0, 1 sequence develops only two Q branches ($Q_2+{}^QP_{21}$ and ${}^PQ_{12}+P_1$) fairly well, while one of the P-head sequences (P_2) is completely absent, and the other P heads (${}^OP_{12}$) are very weak. The 0, 1 sequence on the other hand develops only one Q-head sequence, and this Q-head sequence goes, as in our case, with the Q_2 of the 0, 0 sequence. An interesting explanation of such a situation in CaF is put forward by Harvey(5) and it is clear that the same may be extended to this case also. From all these considerations we believe that we have correctly placed the 0, 1 sequence in the quantum analysis. An isotopic effect would have been an infallible guide in such analysis but unfortunately no such isotopic heads are observed; cadmium possesses several isotopes but the abundance ratio seems to be unfavourable for their detection, at least in the method of excitation of the bands adopted here.

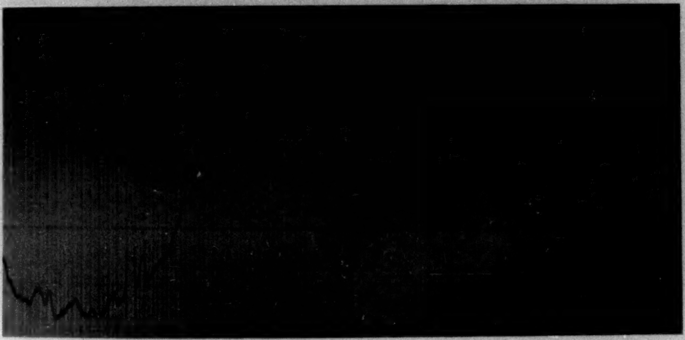


Figure 2.

§ 4. ANALYSIS AND DISCUSSION

In a band system in which a fairly large number of sequences are developed, the method of deriving the vibrational equations is not so tedious. Usually many observations are available of the same frequency difference between adjacent vibrational levels, and the weighted mean or arithmetical mean, as the case may be, of all these values is utilized to obtain the frequency differences between successive vibrational levels in the initial and the final states of the molecule. These are then expressed by a suitable quadratic or cubic formula and the vibrational equation is thence deduced. In cases where, as here, only two sequences are developed, this method is, however, clearly not suitable. We follow Johnson(4) in this case and express each sequence by a polynomial by the rapid least-squares method developed by Birge and Shea(6) and thereby deduce the required vibrational equation by equating the coefficients. Each sequence was first expressed by a polynomial of the third order and so vibrational equations involving cubics were developed, the results being as follows:

Yellow-green system

(i) For R_1 heads:

in 0, 0 sequence, $\nu = 18865\cdot05 - 20\cdot976v'' - 0\cdot124v''^2 + 0\cdot0052v''^3$,
in 0, 1 sequence, $\nu = 18198\cdot58 - 13\cdot23v'' - 0\cdot324v''^2 + 0\cdot0244v''^3$;
hence, for the whole system,
$\nu_{R_1} = 18865\cdot05 + (662\cdot62v' - 3\cdot78v'^2 + 0\cdot067v'^3) - (683\cdot60v'' - 3\cdot65v''^2 + 0\cdot062v''^3)$.

(ii) For R_2 heads:

in 0, 0 sequence, $\nu = 18857\cdot68 - 1841v'' - 1\cdot104v''^2 + 0\cdot064''^3$,
in 0, 1 sequence, $\nu = 18187\cdot78 - 11\cdot991v'' - 0\cdot166v''^2 - 0\cdot0035v''^3$;
hence, for the whole system,
$\nu_{R_2} = 18857\cdot68 + (666\cdot53v' - 3\cdot68v'^2 - 0\cdot313v'^3) - (684\cdot94v'' - 2\cdot58v''^2 - 0\cdot377v''^3)$.

Orange system

For Q_2 heads:
in 0, 0 sequence, $\nu = 16561\cdot05 + 6\cdot389v'' - 0\cdot591v''^2 + 0\cdot055v''^3$,
in 0, 1 sequence, $\nu = 15840\cdot65 + 27\cdot85v'' - 2\cdot82v''^2 + 0\cdot15v''^3$;
hence, for the whole system,
$\nu_{Q_2} = 16561\cdot05 + (710\cdot05v' - 9\cdot615v'^2 + 0\cdot741v'^3) - (703\cdot66v'' - 9\cdot024v''^2 + 0\cdot686v''^3)$.

In view of the fact already mentioned that the accuracy claimed is not very high, it was also thought advisable to use only a quadratic function; the equations thus obtained are:

Yellow-green system

(i) For R_1 heads:

in 0, 0 sequence, $\nu = 18864\cdot75 - 20\cdot98v'' - 0\cdot1174v''^2$,
in 0, 1 sequence, $\nu = 18200\cdot19 - 14\cdot722v'' + 0\cdot0422v''^2$;
hence, for the whole system,
$\nu_{R_1} = 18864\cdot75 + (661\cdot75v' - 3\cdot11v'^2) - (682\cdot43v'' - 3\cdot07v''^2)$.

(ii) For R_2 heads:

in 0, 0 sequence, $\nu = 18859\cdot99 - 22\cdot09v'' - 0\cdot140v''^2$,
in 0, 1 sequence, $\nu = 18187\cdot61 - 11\cdot81v'' - 0\cdot214v''^2$;
hence, for the whole system,
$\nu_{R_2} = 18859\cdot99 + (667\cdot24v' - 5\cdot14v'^2) - (689\cdot33v'' - 4\cdot96v''^2)$.

Orange system

For Q_2 heads:
in 0, 0 sequence, $\nu = 15561\cdot21 + 5\cdot636v'' - 0\cdot1785v''^2$,
in 0, 1 sequence, $\nu = 15856\cdot857 + 17\cdot114v'' - 0\cdot5714v''^2$;
hence, for the whole system,
$\nu_{Q_2} = 16561\cdot21 + (689\cdot62v' - 5\cdot74v'^2) - (692\cdot97v'' - 5\cdot36v''^2)$.

It is found that the cubic expressions on the whole give smaller *O–C* values, but the general equations deduced from them invariably lead to positive values of v^3. This is not a happy situation for they would lead to imaginary values of the heats of dissociation. A similar situation is already met with in the case of the alkaline-earth halides(4). One way of explaining this is that, in such molecules, Q heads may not represent the origins of the bands(5), and the difference $\nu_h - \nu_o$ is not constant throughout the system. Coupled with this are to be taken into consideration the violent perturbations observed in the present molecule. While some of these may really not be perturbations at all, but only errors of observation, we feel convinced that at least the band in the yellow-green system at ν 18741 is a genuine instance. For these reasons we adopt the quadratic equations to represent the band-heads. Reducing these to the form required by the new quantum theory, we have:

Yellow-green system

For R_1 heads:

$$\nu = 18874{\cdot}04 + [664{\cdot}86\,(v'+\tfrac{1}{2}) - 3{\cdot}11\,(v'+\tfrac{1}{2})^2] - [685{\cdot}50\,(v''+\tfrac{1}{2}) - 3{\cdot}07\,(v''+\tfrac{1}{2})^2].$$

For R_2 heads:

$$\nu = 18871{\cdot}1 + [672{\cdot}38\,(v'+\tfrac{1}{2}) - 5{\cdot}14\,(v'+\tfrac{1}{2})^2] - [694{\cdot}29\,(v''+\tfrac{1}{2}) - 4{\cdot}96\,(v''+\tfrac{1}{2})^2].$$

Orange system

For Q_2 heads:

$$\nu = 16558{\cdot}31 + [704{\cdot}36\,(v'+\tfrac{1}{2}) - 5{\cdot}74\,(v'+\tfrac{1}{2})^2] - [698{\cdot}34\,(v''+\tfrac{1}{2}) - 5{\cdot}36\,(v''+\tfrac{1}{2})^2].$$

In these equations we find again many discrepancies, such as are encountered in the case of similar alkaline earth fluorides. We assume that the final level of these systems is the same. The only justification for this is that the final vibration frequencies for the orange system and the yellow-green R_2 heads are nearly equal though the anharmonic constants widely vary. The yellow-green R_1 heads ought really to give roughly the same equations as R_2 heads, but here again the discrepancy —especially in the anharmonic constant—is very noticeable*. The difference however is not surprisingly large if we remember that the R_1 and R_2 branches of a fully resolved band are usually very different in form. These discrepancies can only

---

* To satisfy ourselves that this is not a weighty argument against the analysis, we have calculated the equation for the similar bands of the CaF molecule which also shows a wide discrepancy, especially in the anharmonic constant. Thus the R_2 heads of the green bands of CaF are represented by the following equation (*vide* Johnson)(4):

$$\nu_{\text{head}} = 18888{\cdot}136 + (563{\cdot}388\,v' - 1{\cdot}7160\,v'^2 - 0{\cdot}02245\,v'^3) - \left(583{\cdot}829\,v'' - 1{\cdot}6339\,v''^2 - \begin{Bmatrix} 0{\cdot}02140 \\ 0{\cdot}02241 \end{Bmatrix} v''^3\right).$$

The equations for R_1 heads were calculated using Johnson's data in the same way as he has done, with the following result:

$$\nu_{\text{head}} = 18894{\cdot}652 + (559{\cdot}818\,v' - 1{\cdot}265\,v'^2 + 0{\cdot}1983\,v'^3) - \left(580{\cdot}499\,v'' - 0{\cdot}505\,v''^2 + \begin{Bmatrix} 0{\cdot}1995 \\ 0{\cdot}1991 \end{Bmatrix} v''^3\right).$$

It is true that these equations are not correct in so far as Johnson's $v'\,v''$ numbering is too high by one unit in both v' and v''. But still it is evident that the difference between the coefficients for R_1 and R_2 heads would certainly occur in the corrected equations. Similar discrepancies have also been observed in BeF and other molecules.

be set right by a detailed rotational analysis of the bands, but we believe that the vibrational frequencies derived are of the correct order of magnitude. Under these circumstances, for reasons already mentioned, the data on R_2 heads only were utilized for calculating the heats of dissociation in the case of the yellow-green system. These can be only rough, probably correct to not more than 0·5 V. Before proceeding to discuss this point in detail we should like to draw attention to two more points. (i) The vibrational frequency, which is of the order of 700 cm.$^{-1}$, is higher than for the similar CaF, SrF, and other molecules. That this is of the right magnitude, however, can be seen in table 3 which shows how the frequency of vibration changes as we pass from the hydride to the fluoride in all these cases. The ionization-potential data of the metals concerned also bear this out.

Table 3

Atom	Ionization potential of atom	Hydrides ω_e	Fluorides ω_e	Ratio $\frac{\omega_e \text{ hydride}}{\omega_e \text{ fluoride}}$
Mg	7·61	1493·5	690·8	2·16
Ca	6·09	1316·7	586·7	2·24
Sr	5·67	—	500·1	—
Ba	5·19	—	468·9	—
Cd	8·95	1430·7	692·7	2·07

(ii) The doublet separation for the CdH molecule is about 1000 cm.$^{-1}$ and from analogy we should expect the doublet interval in CdF to be roughly about 900 cm.$^{-1}$; but we find that this is only 75 cm.$^{-1}$. In the 0, 0 sequence of the orange system the P_1 heads are rather diffuse, but the remaining three, Q_1, P_2 and Q_2, are quite sharply developed as can be seen from the enlarged spectrogram, figure 1. Thus there seems to be no doubt about the separation between the Q_1 and Q_2 heads being roughly 75 cm.$^{-1}$. Of course there is the possibility that this is not the deepest $^2\Pi$ state of the molecule, but only a higher excited one. A similar situation was met with in BaH and a lower $^2\Pi$ state was afterwards found with about the right interval. Because of this divergence between the expected and the observed values we have several times tried to see if another arrangement of the heads could be made but without success. The present arrangement is the most straightforward one in such cases and the development of the 0, 0 sequence with its four series cannot be mistaken. We therefore point out this glaring discrepancy and think that very probably this $^2\Pi$ term in the CdF molecule does not arise from the same electronic configuration as the $^2\Pi$ term in CaF and SrF, as will be shown in the discussion on the heats of dissociation. Investigations on the spectrum of the molecule in absorption might throw more light on this point.

§ 5. HEATS OF DISSOCIATION, AND THE STRUCTURE OF CdF

Assuming a linear extrapolation of the vibrational levels one can deduce the heats of dissociation of the CdF molecule in the ground state and the excited states by means of the following well-known formula:

$$D = \frac{\omega_e^2}{4x_e\omega_e \times 8100},$$

where D is the heat of dissociation in volts. The values thus obtained are shown in table 4.

Table 4

Level	D (volts)	Remarks
Ground $^2\Sigma$	$\begin{cases} 3{\cdot}0 \\ 2{\cdot}8 \end{cases}$	From the yellow-green system (R_2 heads) From the orange system
Excited $^2\Sigma$	2·7	From the yellow-green system (R_2 heads)
Excited $^2\Pi$	2·7	

From these data the potential-energy curves in figures 3 and 4, p. 244, were drawn. The course of these curves is drawn qualitatively but with due regard to the direction of degradation of the bands. The difference between the energies of the dissociation products in the final and the excited states is the same, namely 2·0 V., for both band systems. The only possibility of correlating this energy with the energy of excitation of the constituent atoms is the one shown in the diagrams; i.e. the normal state of the CdF molecule arises thus:

$$\text{Cd}\ (5s,\ 5p)\ ^3P + \text{F}\ (2p)^5\ ^2P \rightarrow \text{CdF},\ ^2\Sigma;$$

and the excited states thus:

$$\text{Cd}\ (5s,\ 6s)\ ^3S + \text{F}\ (2p)^5\ ^2P \rightarrow \text{CdF},\ ^2\Sigma \text{ and } ^2\Pi.$$

The Cd atom in its normal state has, in addition to closed shells, two electrons in the $5s$ group which also is completed, and the fluorine atom five electrons in the $2p$ group. In this state of the Cd atom we have only repulsion on account of the closed group and the linkage starts from the excited level of the configuration (sp).

In the case of the alkaline earths Mulliken(7) assumed that linkage starts with the configuration $(s)^2$ whereas Herzberg(3) showed that the linkage probably starts from the first excited term of the metal atom with the configuration (sp). The helium-like configuration $(s)^2$ should give rise only to a repulsive curve of the two-centre system and this indeed seems to be very probable from the paper of Lessheim and Samuel(2) who have shown that if the ground level of these molecules is considered as arising out of the (sp) configuration, then it is possible to explain why the heat of dissociation in some of the excited states in the alkaline-earth halides and in some other molecules is bigger than in the ground state, and in some cases smaller. The p electron of the metal together with a p electron of the halogen are responsible for the chemical linkage. If the p electron which brings about the linkage is excited, we get

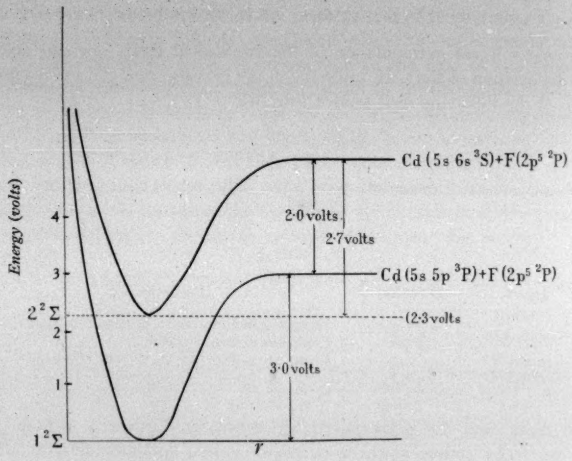

Energy (volts)
4
3
$2\,^2\Sigma$
2
1
$1\,^2\Sigma$
r
Cd (5s 6s ^{3}S)+F(2p^5 ^{2}P)
2·0 volts
2·7 volts
Cd (5s 5p ^{3}P)+F (2p^5 ^{2}P)
(2·3 volts
3·0 volts

Figure 3. Yellow-green CdF bands (degraded to red).

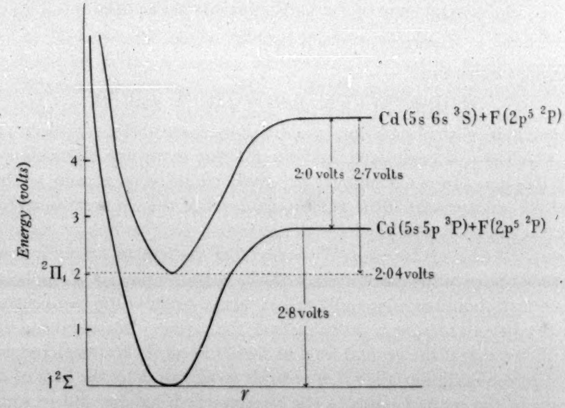

Energy (volts)
4
3
$^2\Pi_1$ 2
1
$1^2\Sigma$
r
Cd (5s 6s ^{3}S)+F (2p^5 ^{2}P)
2·0 volts
2·7 volts
Cd (5s 5p ^{3}P)+F (2p^5 ^{2}P)
2·04 volts
2·8 volts

Figure 4. Orange CdF bands (degraded to violet).

a molecular electronic state which has a decreased heat of dissociation and which dissociates adiabatically, yielding a metal in a normal excited state. The *s* electron of the metal does not take part in the linkage but only disturbs it. Therefore if this *s* electron in the molecule is now excited, the heat of dissociation increases in the excited level and the term dissociates adiabatically yielding a metal atom in which two electrons are excited, i.e. a metal atom in an anomalous excited state.

For these reasons the ground state of the CdF molecule arises from the configuration Cd $(5s, 5p)\,{}^3P$, since the configuration Cd $(5s)^2$ gives rise to a repulsive curve only. It can be shown that the spectra of the alkaline-earth oxides, and Hund's theory of their crystal lattice, also lend support to this view$^{(8)}$. In this case, however, we cannot be certain of the quantum numbers of the united atom and we have recourse to the configuration in the analogous case of BeF. Here we obtain from the configurations of the metal and fluorine the following configuration for the level of the molecule according to Weizel's correlation scheme:

$$^2\Sigma\ :\ 2s\sigma^2\,(2s),\ 2p\pi^4\,(2p),\ 3p\sigma\,(2s),\ 3d\sigma^2\,(2p).$$

Since the dissociation energy decreases in the excited terms observed in CdF, we will consider those configurations of the excited molecule in which the excitation is due to one of the electrons contributing to the linkage; this electron will be either in π^4 or σ^2 group. The excitation will bring it to either the next σ group or the next π group. There are two possible σ groups, $3s\sigma\,(3s)$ and $4p\sigma\,(3s)$, and two possible π groups, $3p\pi\,(3p)$ and $3d\pi\,(2p)$. Since the dissociation energy indicates the electronic configuration $(5s, 6s)$ of Cd, the π group containing p electrons of the separated atoms may be excluded. We therefore derive the following configurations for the excited levels of BeF:

$$^2\Pi\ :\ 2s\sigma^2\,(2s),\ 2p\pi^3\,(2p),\ 3p\sigma\,(2s),\ 3d\sigma^2\,(2p),\ 3s\sigma\,(3s)$$

and

$$^2\Sigma\ :\ 2s\sigma^2\,(2s),\ 2p\pi^4\,(2p),\ 3p\sigma\,(3s),\ 3d\sigma\,(2p),\ 3s\sigma\,(3s);$$

in which the $4p\sigma\,(3s)$ may replace the $3s\sigma\,(3s)$ and the $^2\Pi$ states may be inverted. Both states will have less energy of dissociation. Neglecting now the quantum numbers of the united atom, we suggest therefore the following configurations for the three levels of CdF:

$^2\Sigma$ ground state : $\sigma^2\,(s),\ \pi^4\,(p),\ \sigma\,(s),\ \sigma^2\,(p).$
$^2\Pi$ term : $\sigma^2\,(s),\ \pi^3\,(p),\ \sigma\,(s),\ \sigma^2\,(p),\ \sigma\,(s).$
$^2\Sigma$ excited state : $\sigma^2\,(s),\ \pi^4\,(p),\ \sigma\,(s),\ \sigma\,(p),\ \sigma\,(s).$

Of these, the excited $^2\Sigma$ state is similar to the excited $^2\Sigma$ state observed in the case of CaF and SrF molecules. The difference in the energies of the Cd atom in the $(5s, 6s)\,{}^3S$ and $(5s, 5p)\,{}^3P$ states is 2·4 V.$^{(9)}$. In view of the errors of observation, and for reasons previously explained, we think that the value 2, 0 obtained from the bands is in fair agreement with this value. The excited $^2\Pi$ state cannot be compared to the $^2\Pi$ state of the alkaline-earth halides. In the case of CaF and other such molecules, the $^2\Pi$ state has a greater dissociation energy than in the ground state and arises out of the anomalous term. In the case of the CdF molecule, the excited $^2\Pi$ has a smaller dissociation energy and arises out of the same terms that give the excited $^2\Sigma$ state. The reason is that the metals Zn, Cd and Hg of the sub-group of the periodic system have a much higher ionization potential than the alkaline earths. Therefore the anomalous terms of these atoms are also much higher than for the atoms of the main group. The same is already the case in BeF. The anomalous term $(5p^2)\,^3P$ in the Cd atom lies about 5·4 V. above the $(5s, 5p)\,^3P$ term, while in Ca this anomalous term is 2·9 V. above the $(4s, 4p)\,^3P$ term(9). Since the dissociation energies of the two molecules in their ground states are nearly the same, we assume that the dissociation energy of the CdF molecule which arises out of this configuration in the Cd atom is also about the same as that for the CaF molecule in a similar state, namely 4·8 V. Calculation will show that a transition from such a $^2\Pi$ state to the ground $^2\Sigma$ state should give rise to bands near about λ3100. The plates do not show any trace of such a system in this region under the conditions of the present investigation.

REFERENCES

(1) Jevons. *Report on Band-Spectra of Diatomic Molecules* (Phys. Soc. 1932). Also for general reference.
(2) Lessheim and Samuel. *Z. Phys.* **84**, 637 (1933).
(3) Herzberg. *Z. Phys.* **75**, 601 (1930).
(4) Johnson. *Proc. roy. Soc.* A, **122**, 161 (1929).
(5) Harvey. *Proc. roy. Soc.* A, **133**, 336 (1931).
(6) Birge and Shea. *Univ. Calif. Publ. Math.* **2**, 67 (1927).
(7) Mulliken. *Rev. mod. Phys.* **4**, 1–86 (1932).
(8) R. Samuel. "Report on Absorption Spectra," *Proc. Ind. Acad. Sci.* **1** (in press).
(9) Bacher and Goudsmit. *Atomic Energy States* (McGraw-Hill, 1932).

535.338.4:546.182.6

ROTATIONAL ANALYSIS OF THE ULTRA-VIOLET BANDS OF PHOSPHORUS MONOXIDE

By A. K. SEN GUPTA

Communicated by Prof. P. N. Ghosh, October 16, 1934. *Read in title December* 7, 1934.

ABSTRACT. The structure of (0, 0), (0, 1) and (1, 0) bands of the ultra-violet system of the PO molecule has been analysed. The bands are due to a $^2\Sigma \rightarrow {}^2\Pi$ transition. Each band consists of six main branches P_2, Q_2, R_2, and P_1, Q_1, R_1 and two satellite branches $^OP_{12}$ and $^SR_{21}$. The potential-energy curves for both the upper and lower states of the system have been drawn on the basis of the equations given by Morse as well as by Rydberg. From these curves the Condon parabola for the intensity of the bands in the system has been obtained. The relevant molecular constants have been obtained. The products of dissociation in the different states are discussed.

§ 1. INTRODUCTION

A VIBRATIONAL quantum analysis of the ultra-violet bands of phosphorus monoxide has been recently published by Ghosh and Ball⁽¹⁾. The marked resemblance in appearance which these bands bear to the γ bands of NO led these authors to suggest that they might be due to a $^2\Sigma \rightarrow {}^2\Pi$ transition.

In the present investigation, a rotational analysis of the (0, 0), (0, 1) and (1, 0) bands has been made. The analysis confirms the idea that they are of the $^2\Sigma \rightarrow {}^2\Pi$ type. Each band is found to consist of eight branches, viz., $^OP_{12}$, P_2, Q_2, R_2 and P_1, Q_1, R_1, $^SR_{21}$.

§ 2. SPECTROGRAMS

Using as a source the flame of carbon arc fed from 220-volt d.c. circuit with P_2O_5 placed in the lower (positive) electrode and with a current of 3–4 A., the bands have been photographed in the 1st order of a 21-ft. concave grating recently set up in the laboratory on a Paschen mounting and having a dispersion of about 1·22 A./mm. in the first order in the neighbourhood of λ 2500. The time of exposure was about 15 hours. The usual standard iron arc was used as a comparison spectrum. For measurements a Gaertner precision comparator was used. The wave-numbers given in tables 2, 3 and 4 are correct to $\pm 0{\cdot}2$ cm.$^{-1}$

§ 3. ROTATIONAL ANALYSIS

As the bands were excited at a high temperature, the different series were long and there was considerable overlapping of series lines from one band to another. Many of the observed lines were blends. The bands (0, 0), (0, 1) and (1, 0) were selected for measurement of structure lines in view of their comparatively strong intensity and minimum overlapping due to the lines of the neighbouring bands. The choice of these bands also enabled us to take full advantage of the combination principle.

Each band is found to consist of eight branches, four of them forming heads. This shows that the band system is due to a $^2\Sigma \rightleftarrows {}^2\Pi$ transition, with negligible spin doubling in the $^2\Sigma$ state. A more definite idea about the nature of transition could not be derived from the criterion of missing lines, for lines of low J values are not sufficiently intense for measurement. The intensity of the different branch lines for moderate J values is, however, in agreement with what one would expect in the case of a $^2\Sigma \rightarrow {}^2\Pi$ transition. In fact, the analysis of each individual band was started on this assumption so that the four heads were attributed to $^OP_{12}$, P_2, P_1 and Q_1 branches. Members of the P, Q, and R series with common values of rotation quantum number of the lower energy state were selected with the aid of the relation

$$R(J)-Q(J) \fallingdotseq Q(J+1)-P(J+1),$$

and those of the $^OP_{12}$ and $^SR_{21}$ series from the relations

$$R_1(J)-P_1(J)=R_2(J)-{}^OP_{12}(J)$$

and $$R_2(J)-P_2(J)={}^SR_{21}(J)-Q_1(J).$$

§4. EVALUATION OF THE CONSTANTS

The upper $^2\Sigma$ *state.* The upper-state combinations are defined by

$$R_1(J)-P_1(J)=\Delta_2T_1'(J), \ldots J=K+\tfrac{1}{2}$$

and $$R_2(J)-P_2(J)=\Delta_2T_2'(J), \ldots J=K-\tfrac{1}{2}.$$

Within the limits of experimental error, these differences for (0, 0) and (0, 1) bands are equal in magnitude. This indicates that the two bands have a common upper state and that the spin doubling of the $^2\Sigma$ state is negligible. The values of B_v' and D_v' for the upper state were calculated by the method of successive approximation from the mean of these differences in accordance with the following relation:

$$\Delta_2T'(K)=B_v'(4K+2)+D_v'(8K^3+12K^2+12K+4).$$

The lower $^2\Pi$ *state.* The lower-state combinations are defined by

$$R_1(J-1)-P_1(J+1)=\Delta_2T_1''(J), \text{ for the } {}^2\Pi_{\frac{1}{2}} \text{ sub-state,}$$

and $$R_2(J-1)-P_2(J+1)=\Delta_2T_2''(J), \text{ for the } {}^2\Pi_{1\frac{1}{2}} \text{ sub-state.}$$

The mutual agreement between these differences in each sub-state for (0, 0) and (1, 0) bands indicates that the two bands have a common lower state. The values of B_v'' and D_v'' were calculated from the relation

$$\Delta_2T''(J)=B_v''(4J+2)+D_v''(8J^3+12J^2+12J+4).$$

The values of the molecular constants are given in table 1.

Table 1. Constants of the PO molecule.

Upper $^2\Sigma$ state	Lower $^2\Pi_{\frac{1}{2}}$ sub-state	Lower $^2\Pi_{1\frac{1}{2}}$ sub-state
$B_e'=0{\cdot}8121$ cm.$^{-1}$	$B_e''=0{\cdot}7613$ cm.$^{-1}$	$B_e''=0{\cdot}7645$ cm.$^{-1}$
$B_0'=0{\cdot}8093$ cm.$^{-1}$	$B_0''=0{\cdot}7585$ cm.$^{-1}$	$B_0''=0{\cdot}7617$ cm.$^{-1}$
$B_1'=0{\cdot}8037$ cm.$^{-1}$	$B_1''=0{\cdot}7530$ cm.$^{-1}$	$B_1''=0{\cdot}7652$ cm.$^{-1}$
$\alpha'=0{\cdot}0056$ cm.$^{-1}$	$\alpha''=0{\cdot}0055$ cm.$^{-1}$	$\alpha''=0{\cdot}0055$ cm.$^{-1}$
$D_e'=-1{\cdot}096\times10^{-6}$ cm.$^{-1}$	$D_e''=-1{\cdot}151\times10^{-6}$ cm.$^{-1}$	$D_e''=-1{\cdot}167\times10^{-6}$ cm.$^{-1}$
$r_e'=1{\cdot}402\times10^{-8}$ cm.	$r_e''=1{\cdot}446\times10^{-8}$ cm.	$r_e''=1{\cdot}443\times10^{-8}$ cm.
$I'=34{\cdot}109\times10^{-40}$ gm. cm.^{2}	$I''=36{\cdot}385\times10^{-40}$ gm. cm.^{2}	$I''=36{\cdot}233\times10^{-40}$ gm. cm.^{2}

§ 5. Δ-TYPE DOUBLING

The combination defects in the (0, 1), (0, 0) and (1, 0) bands were computed for each value of J in accordance with the following relation:

$$2\Delta\nu_{dc}\,(J+\tfrac{1}{2})=\{Q\,(J)-P\,(J+1)\}-\{R\,(J)-Q\,(J+1)\}.$$

The combination defects obtained from R_1, P_1, and Q_1 branches are negative and increase almost linearly with $(J+\frac{1}{2})$ for moderate J values, while those obtained from R_2, P_2 and Q_2 branches are positive and rather smaller in magnitude. This shows that R_1, P_1 and Q_1 branches are to be associated with $^2\Pi_{\frac{1}{2}}$, and R_2, P_2 and Q_2 branches with $^2\Pi_{1\frac{1}{2}}$ sub-states, indicating that $^2\Pi$ state is normal. The nature of the combination defects points further to the fact that the coupling is intermediate between case *a* and case *b* and is nearer to case *a*. This is also supported by the fact that the satellite branches $^OP_{12}$ and $^SR_{21}$ extend to fairly large rotational quantum numbers.

§ 6. POTENTIAL-ENERGY CURVES AND INTENSITY DISTRIBUTION IN THE BAND SYSTEM

Potential-energy curves derived by means of both Morse⁽²⁾ and Rydberg⁽³⁾ equations are given in figures 1 and 2. Using the Franck-Condon theory of probability of transition, the Condon parabola for the intensity-distribution in the band system is drawn in figure 3 as deduced from them; *M* indicates the parabola derived from the Morse curves and *R* that derived from those of Rydberg. The Condon parabola *M* seems to be in good agreement with the visual estimation of the intensities of the band-heads, but the parabola *R* fits still better.

§ 7. PRODUCTS OF DISSOCIATION

An enquiry into the products of dissociation of the PO molecule seems to be of interest in view of the conflicting opinions regarding their identity in the case of the NO molecule. Birge and Sponer and others were of opinion that the latter molecule in its different excited states is dissociated into an excited oxygen atom and a normal nitrogen atom. Hence for the homologous PO molecule one would naturally expect that the value of the atomic excitation energy E_a as calculated from its band system should be identical with that obtained from the γ bands of NO. From a similar consideration Lessheim and Samuel⁽⁴⁾ calculated the values of E_a in the case of NO and PO molecules and found them to be of distinctly different magnitude. This led these authors to conclude that the two oxide molecules in their excited states are dissociated into a normal oxygen atom and an excited nitrogen or phosphorus atom. For the PO molecule the heats of dissociation of its different states were calculated by them from the earlier data. In view of the fact that new data are now available from the recent vibrational analysis of Ghosh and Ball, we are now in a position to decide more definitely the identity of the products of dissociation.

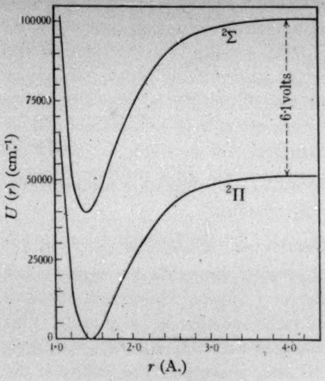


Figure 1. Potential-energy curves by Morse's function.

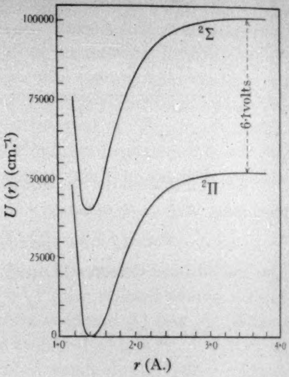


Figure 2. Potential-energy curves by Rydberg's function.

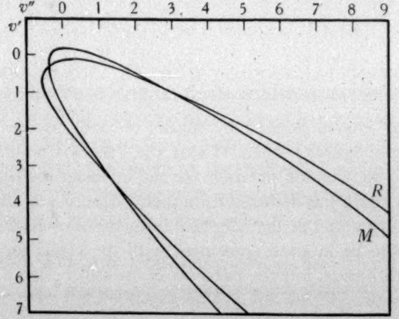


Figure 3. Intensity curves, *M* from figure 1, *R* from figure 2.

From these recent data the value of E_a is calculated to be 6·1 V. and is thus widely different from the atomic excitation energy of the oxygen atom corresponding to the $^5S \rightarrow {}^3P$ transition. It is, however, in fair agreement, within the limits of extrapolation error, with the E_a value of phosphorus, if we assume that the $^2\Pi$ state of the PO molecule is derived from normal P (^4S) and O (^3P) atoms and its upper $^2\Sigma$ state from excited P (^4P) and normal O (^3P) atoms. From the data of energy states published by Bacher and Goudsmit (5) the difference of energy values between a P atom in the normal ($3s^2 3p^3$) ^4S state and that in an excited ($3s^2 3p^2 4s$) ^4P state is 6·9 V. This result, therefore, appears to be in agreement with the view of Lessheim and Samuel.

There is yet another possibility that the products of dissociation of the PO molecule in its normal and excited states may consist respectively of a P atom in ($3s^2 3p^3$) ^2D and ($3s^2 3p^2 4s$) ^2P states, since the difference of energy values between these two atomic states is 5·7 V. This indicates that in the normal state the molecule is dissociated into a normal O atom and a metastable P atom. A similar possibility may be reckoned also in the case of the NO molecule, since the atomic excitation energy derived from the γ bands is 9·4 V., whereas the differences between the energy states of N atom corresponding to $^4P \rightarrow {}^4S$ and $^2P \rightarrow {}^2D$ are 10·3 and 8·3 V. respectively. From the uncertainties involved in the determination of dissociation energies by the method of extrapolation, it is difficult to decide between the two alternatives.

§ 8. DISCUSSION OF RESULTS

A check on the correctness and the analysis was afforded by a rule due to R. T. Birge (6). According to this rule the quantity $2x_e B_e/\alpha$ is approximately equal to $1{\cdot}4 \pm 0{\cdot}2$. For these bands the substitution of the proper values showed that $2x_e' B_e'/\alpha' = 1{\cdot}59$ and $2x_e'' B_e''/\alpha'' = 1{\cdot}47$.

An additional check is given by the approximate relation that for a molecule composed of two atoms of nearly equal mass, the quantity $\omega_e r_e^3$ is approximately equal to 3000×10^{-24} cm.² This condition is satisfied since we have

$$\omega_e' r_e'^3 = 3800 \times 10^{-24} \text{ cm}^2 \quad \text{and} \quad \omega_e'' r_e''^3 = 3700 \times 10^{-24} \text{ cm}^2$$

It is to be expected that for PO the value of $\omega_e r_e^3$ will be greater than 3000×10^{-24} cm.² since the masses of the two atoms are rather unequal.

§ 9. ACKNOWLEDGMENT

With great pleasure I acknowledge my indebtedness to Prof. Dr P. N. Ghosh for his continual guidance and helpful suggestions during the course of the investigation.

REFERENCES

(1) P. N. GHOSH and G. N. BALL. *Z. Phys.* **71**, 362 (1931).
(2) P. M. MORSE. *Phys. Rev.* **34**, 57 (1929).
(3) R. RYDBERG. *Z. Phys.* **73**, 376 (1931–2).
(4) H. LESSHEIM and R. SAMUEL. *Z. Phys.* **84**, 637 (1933).
(5) R. F. BACHER and S. GOUDSMIT. *Atomic Energy States* (1932).
(6) R. T. BIRGE. *Phys. Rev.* **31**, 919 (1927).

Table 2. (0, 0) band at λ 2478

$^2\Sigma \rightarrow {}^2\Pi_{1\frac{1}{2}}$

J''	$^{O}P_{12}$	P_2	Q_2	R_2	$\Delta_2 T_2'(J)$	$\Delta_2 T_2''(J)$
1½	—	—	40376·83	—	—	—
2½	—	—	377·91	—	—	—
3½	—	—	378·95	—	—	—
4½	40365·74	—	380·25	—	—	—
5½	363·95	—	381·62	40392·97	—	18·34
6½	362·06	40371·61	383·03	395·93	24·32	21·37
7½	360·36	371·61	384·48	399·10	27·49	24·32
8½	358·57	371·61	386·29	402·45	30·84	27·49
9½	357·10	371·61	387·82	405·63	34·02	—
10½	355·80	—	389·71	408·08	—	33·43
11½	354·34	372·20	391·60	412·73	40·53	36·49
12½	353·13	372·59	393·63	416·33	43·74	39·65
13½	352·10	373·08	395·83	420·99	46·91	42·60
14½	350·92	373·73	397·87	423·68	49·95	45·71
15½	349·97	374·28	400·12	427·60	53·32	48·65
16½	349·12	375·03	402·55	431·69	56·66	51·75
17½	348·47	375·85	405·06	435·72	59·87	54·86
18½	347·82	376·83	407·61	439·89	63·06	57·91
19½	347·23	377·81	410·12	444·12	66·31	60·94
20½	346·68	378·95	413·06	448·62	69·67	63·87
21½	346·36	380·25	415·84	453·02	72·77	66·90
22½	—	381·72	418·78	457·54	75·82	69·99
23½	345·87	383·03	421·89	462·20	79·17	72·96
24½	345·87	384·58	424·99	466·96	82·38	75·91
25½	345·87	386·29	428·35	471·95	85·66	79·04
26½	345·87	387·92	431·53	476·71	88·79	82·24
27½	—	389·71	434·96	481·87	92·16	85·12
28½	346·36	391·59	438·56	486·97	95·38	88·14
29½	346·88	393·73	442·25	492·12	98·39	91·14
30½	347·43	395·83	445·86	497·52	101·69	94·08
31½	347·98	398·04	449·68	502·88	104·84	97·20
32½	348·68	400·32	453·61	508·46	108·14	100·19
33½	349·45	402·69	457·59	514·09	111·40	103·23
34½	350·43	405·23	461·80	519·78	114·55	106·33
35½	351·30	407·76	465·89	525·60	117·84	109·33
36½	352·38	410·45	470·15	531·35	120·90	112·37
37½	353·68	413·23	474·49	537·27	124·04	115·18
38½	354·92	416·17	479·08	543·51	127·34	118·16
39½	356·28	419·11	483·59	549·60	130·49	121·30
40½	357·76	422·21	488·38	555·84	133·63	124·20
41½	359·38	425·40	493·02	562·35	136·95	127·24
42½	361·02	428·60	498·06	568·68	140·08	130·33
43½	362·81	432·02	502·78	575·27	143·25	133·39
44½	364·68	435·29	507·89	581·86	146·57	136·38
45½	366·56	438·89	513·06	588·70	149·81	139·21
46½	368·61	442·65	518·45	595·53	152·88	142·12
47½	370·79	446·58	523·81	602·46	155·88	145·29
48½	373·08	450·24	528·95	609·47	159·23	148·36
49½	375·52	454·10	534·74	616·55	162·45	151·11
50½	377·91	458·36	540·32	623·74	165·38	154·03
51½	380·52	462·52	546·05	631·17	168·65	—
52½	383·13	—	551·74	638·44	—	—
53½	—	—	557·66	646·04	—	—
54½	—	—	563·91	653·64	—	—
55½	—	—	569·83	661·41	—	—
56½	—	—	576·09	669·03	—	—
57½	—	—	582·35	677·13	—	—
58½	—	—	588·70	—	—	—
59½	—	—	595·37	—	—	—

Table 2 (*cont.*)

$^2\Sigma \rightarrow {}^2\Pi_{\frac{1}{2}}$

J''	P_1	Q_1	R_1	$^SR_{21}$	$\Delta_2 T_1'(J)$	$\Delta_2 T_1''(J)$
3½	40590·42	—	40601·63	40609·72	11·21	—
4½	588·29	—	602·79	612·68	14·50	15·16
5½	586·47	—	604·27	615·65	17·80	17·97
6½	584·82	40594·38	605·76	618·79	20·94	21·26
7½	583·01	594·38	607·41	621·89	24·40	24·41
8½	581·36	594·38	609·06	625·23	27·70	27·36
9½	580·05	—	610·70	628·53	30·65	30·33
10½	578·73	—	612·68	632·00	33·95	33·29
11½	577·41	595·20	614·66	635·55	37·25	36·43
12½	576·25	595·63	616·65	639·43	40·40	39·39
13½	575·27	596·19	618·89	643·07	43·62	42·37
14½	574·28	596·89	621·16	647·03	46·88	45·43
15½	573·46	597·64	623·59	651·13	50·13	48·48
16½	572·68	598·50	626·05	655·13	53·37	51·53
17½	572·06	599·45	628·63	659·26	56·57	54·57
18½	571·48	600·54	631·34	663·64	59·86	—
19½	—	601·70	633·98	668·03	—	60·52
20½	—	602·92	637·05	672·45	—	63·65
21½	570·33	604·40	639·93	677·13	69·60	66·72
22½	570·33	605·76	643·00	681·79	72·67	69·60
23½	570·33	607·41	646·30	686·56	75·97	72·67
24½	570·33	609·12	649·57	691·59	79·24	—
25½	—	610·83	652·98	696·50	—	78·75
26½	570·82	612·80	656·45	701·67	85·63	81·75
27½	571·23	614·73	660·09	706·86	88·86	84·80
28½	571·65	616·87	663·74	712·24	92·09	87·79
29½	572·30	619·02	667·53	717·55	95·23	90·61
30½	573·13	621·43	671·44	723·19	98·31	93·66
31½	573·87	623·74	675·47	728·66	101·60	96·67
32½	574·77	626·30	679·60	734·47	104·83	99·84
33½	575·76	628·82	683·84	740·21	108·08	102·92
34½	576·75	631·57	688·12	746·15	111·37	105·77
35½	578·07	634·37	692·52	752·24	114·45	108·73
36½	579·39	637·18	697·10	758·39	117·71	111·82
37½	580·70	640·30	701·70	764·49	121·00	114·75
38½	582·35	643·36	706·37	770·92	124·02	117·70
39½	584·00	646·67	711·25	777·34	127·25	120·73
40½	585·64	649·87	716·22	783·74	130·58	123·79
41½	587·46	653·44	721·18	790·48	133·72	126·79
42½	589·43	657·83	726·40	797·14	136·97	129·75
43½	591·43	660·55	731·55	804·03	140·12	132·76
44½	593·64	664·25	736·89	810·94	143·25	135·69
45½	595·86	668·06	742·38	817·96	146·52	138·62
46½	598·27	672·05	747·92	825·10	149·75	141·64
47½	600·74	676·13	753·50	832·23	152·76	144·68
48½	603·24	680·16	759·28	839·70	156·04	147·64
49½	605·86	684·56	765·01	846·97	159·15	150·55
50½	608·73	688·85	771·02	—	162·29	153·48
51½	611·63	693·38	777·01	—	165·38	156·52
52½	614·60	697·89	783·26	—	168·66	159·38
53½	617·63	702·63	789·52	—	171·89	162·33
54½	620·93	707·34	795·91	—	174·98	165·35
55½	624·17	712·27	802·30	—	178·13	168·45
56½	627·46	717·35	808·79	—	181·33	171·13
57½	631·17	722·33	—	—	—	174·15
58½	634·64	727·63	—	—	—	—
59½	638·44	732·90	—	—	—	—
60½	642·40	738·25	—	—	—	—

Table 3. (0, 1) band at λ 2555

$^2\Sigma \rightarrow {}^2\Pi_{1\frac{1}{2}}$

J''	${}^{O}P_{12}$	P_2	Q_2	R_2	$\Delta_2T_2'(J)$	$\Delta_2T_2''(J)$
1½	—	—	39154·34	—	—	—
2½	—	—	154·48	—	—	—
3½	—	—	156·48	—	—	—
4½	39143·30	—	157·71	—	—	—
5½	141·46	—	159·24	39170·60	—	—
6½	139·78	39149·43	160·78	173·62	24·19	21·07
7½	138·09	149·43	162·31	176·89	27·46	24·19
8½	136·56	149·43	164·05	180·19	30·76	—
9½	135·18	—	165·84	183·72	—	30·15
10½	133·80	150·04	167·89	187·17	37·13	33·22
11½	132·73	150·50	169·82	190·84	40·34	36·21
12½	131·65	150·96	172·00	194·63	43·67	39·26
13½	130·74	151·58	174·28	198·54	46·96	42·28
14½	129·74	152·35	176·58	202·55	50·20	45·43
15½	128·74	153·11	179·04	206·50	53·39	48·37
16½	128·13	154·18	181·64	210·69	56·51	51·40
17½	127·72	155·10	184·22	215·00	59·90	54·36
18½	127·21	156·33	187·07	219·41	63·08	57·44
19½	126·91	157·56	189·94	223·87	66·31	60·47
20½	—	158·94	192·93	228·49	69·55	63·40
21½	126·45	160·47	196·08	233·13	72·66	66·49
22½	126·45	162·00	199·14	237·93	75·93	69·44
23½	126·45	163·69	202·55	242·86	79·17	72·40
24½	—	165·53	205·90	247·89	82·36	75·41
25½	126·91	167·45	209·46	253·11	85·66	78·50
26½	127·47	169·39	213·05	258·27	89·88	81·59
27½	127·98	171·52	216·77	263·55	92·03	84·45
28½	128·74	173·82	220·61	269·06	95·24	87·50
29½	129·41	176·05	224·60	274·62	98·57	90·49
30½	130·17	178·57	228·69	280·28	101·71	93·51
31½	131·04	181·11	232·77	286·00	104·89	96·46
32½	132·11	183·82	237·16	291·91	108·09	99·44
33½	133·34	186·56	241·37	297·77	111·21	102·58
34½	134·51	189·33	245·88	303·87	114·54	105·30
35½	135·94	192·47	250·51	310·13	117·66	108·40
36½	137·32	195·47	255·18	316·37	120·90	111·47
37½	138·96	198·70	259·97	322·81	124·11	114·44
38½	140·69	201·93	264·90	329·31	127·38	117·50
39½	142·53	205·31	269·88	335·81	130·50	120·31
40½	144·45	209·00	275·00	342·62	133·62	123·43
41½	146·36	212·58	280·13	349·36	136·98	126·39
42½	148·66	216·23	285·52	356·25	140·02	129·29
43½	150·96	220·07	290·92	363·30	143·23	132·17
44½	153·19	224·08	296·54	370·56	146·48	135·18
45½	155·77	228·12	302·25	377·74	149·62	138·25
46½	158·32	232·31	307·97	385·14	152·83	141·19
47½	161·00	236·55	313·77	392·57	156·02	144·18
48½	163·69	240·96	319·87	400·14	159·18	147·09
49½	166·61	245·48	325·90	407·75	162·27	149·96
50½	169·68	250·08	332·09	415·50	165·42	153·03
51½	172·82	254·72	338·28	423·37	168·65	155·92
52½	176·05	259·58	344·84	431·53	171·95	158·87
53½	179·34	264·54	351·35	439·62	175·12	162·00
54½	182·88	269·53	358·00	447·71	178·18	164·85
55½	186·41	274·77	364·66	455·96	181·19	167·69
56½	190·25	280·02	371·54	464·38	184·36	170·54
57½	193·94	285·42	378·49	473·02	187·60	173·56
58½	197·93	290·82	385·57	481·57	190·75	176·48
59½	201·77	296·54	392·85	—	—	179·57
60½	—	302·10	—	—	—	—

Table 3 (*cont.*)

$^2\Sigma \rightarrow {}^2\Pi_{\frac{1}{2}}$

J''	P_1	Q_1	R_1	$^{S}R_{21}$	$\Delta_2 T_1'(J)$	$\Delta_2 T_1''(J)$
3½	39370·35	—	39381·75	39389·82	11·40	—
4½	368·49	—	383·07	392·75	14·58	15·20
5½	366·55	—	384·46	396·11	17·91	17·99
6½	365·08	—	386·06	398·90	20·98	20·93
7½	363·53	39374·79	387·70	402·24	24·17	24·08
8½	361·98	374·79	389·51	405·53	27·54	27·11
9½	360·59	375·24	391·45	409·20	30·86	30·12
10½	359·39	375·52	393·31	412·73	33·92	33·26
11½	358·19	375·93	395·43	416·40	37·24	36·13
12½	357·18	376·60	397·65	420·30	40·47	39·18
13½	356·25	377·23	399·93	424·23	43·68	42·17
14½	355·48	378·15	402·40	428·26	46·92	45·08
15½	354·85	379·04	404·96	432·46	50·11	—
16½	—	380·14	407·65	436·66	—	51·16
17½	353·80	381·26	410·40	441·02	56·60	54·11
18½	353·54	382·53	413·35	445·53	59·81	—
19½	—	383·85	416·30	450·25	—	60·36
20½	352·99	385·37	419·41	454·87	66·42	63·31
21½	352·99	387·06	422·67	459·85	69·68	—
22½	—	388·81	426·09	464·84	—	69·13
23½	353·54	390·70	429·51	469·77	75·97	72·29
24½	353·80	392·54	433·08	475·02	79·28	—
25½	—	394·61	436·66	480·27	—	78·23
26½	354·85	396·78	440·50	485·78	85·65	81·18
27½	355·48	399·11	444·44	491·19	88·96	84·25
28½	356·25	401·49	448·33	496·80	92·08	87·16
29½	357·28	403·98	452·54	502·63	95·26	90·14
30½	358·19	406·65	456·74	508·40	98·55	93·15
31½	359·39	409·38	461·10	514·44	101·71	96·05
32½	360·69	412·26	465·54	520·54	104·85	99·02
33½	362·08	415·21	470·14	526·55	108·06	102·01
34½	363·53	418·37	474·89	532·88	111·36	104·96
35½	365·18	421·56	479·69	—	114·51	107·95
36½	366·94	424·85	484·63	545·86	117·69	110·97
37½	368·72	428·32	489·70	552·43	120·98	113·97
38½	370·66	431·84	494·81	559·16	124·15	116·95
39½	372·75	435·48	500·08	566·12	127·33	119·96
40½	374·85	439·17	505·39	573·02	130·54	122·91
41½	377·17	443·15	510·92	580·20	133·75	125·77
42½	379·62	447·12	516·57	587·35	136·95	128·76
43½	382·06	451·25	522·25	—	140·19	131·87
44½	384·70	455·39	528·04	—	143·34	134·78
45½	385·47	459·75	534·00	—	146·43	137·68
46½	390·36	464·18	540·03	—	149·67	140·59
47½	393·31	468·78	546·20	—	153·89	143·57
48½	396·46	473·49	552·53	—	156·07	146·53
49½	399·67	478·35	558·84	—	159·17	149·44
50½	403·09	483·23	565·38	—	162·29	152·34
51½	406·51	488·23	571·99	—	165·48	155·35
52½	410·03	493·33	578·73	—	168·70	158·33
53½	413·66	498·65	585·58	—	171·92	161·19
54½	417·54	504·06	—	—	—	164·15
55½	421·43	509·66	—	—	—	—
56½	425·47	515·12	—	—	—	—
57½	429·51	520·90	—	—	—	—
58½	433·86	526·69	—	—	—	—
59½	438·25	532·68	—	—	—	—
60½	442·57	538·72	—	—	—	—

Table 4. (1, 0) band at λ 2396

$^2\Sigma \rightarrow {}^2\Pi_{1\frac{1}{2}}$

J''	$^{O}P_{12}$	P_2	Q_2	R_2	$\Delta_2 T_2'(J)$	$\Delta_2 T_2''(J)$
1½	—	—	41753·66	—	—	—
2½	—	—	754·70	—	—	—
3½	41744·59	—	755·75	—	—	—
4½	742·42	41748·96	757·07	41766·74	17·78	—
5½	740·58	748·59	758·27	769·53	20·94	—
6½	738·74	—	759·58	772·49	—	21·45
7½	736·85	748·08	760·98	775·46	27·38	21·41
8½	735·27	748·08	762·55	778·60	30·52	27·38
9½	733·61	748·08	764·12	781·87	33·79	30·52
10½	732·09	748·08	765·86	785·06	36·98	—
11½	730·74	—	767·69	788·55	—	36·47
12½	729·30	748·59	769·46	791·96	43·37	39·59
13½	728·04	748·96	771·47	795·44	46·48	42·58
14½	726·90	749·38	773·37	799·12	49·74	45·62
15½	725·77	749·82	775·60	802·88	53·06	48·69
16½	724·72	750·43	777·70	806·56	56·13	51·84
17½	723·77	751·10	780·01	810·48	59·38	54·64
18½	722·99	751·92	782·42	814·47	62·55	57·87
19½	722·12	752·61	784·71	818·49	65·88	60·91
20½	721·46	753·56	787·39	822·65	69·09	63·86
21½	720·90	754·63	789·93	826·84	72·21	66·90
22½	720·38	755·75	792·67	831·05	75·30	69·97
23½	720·03	756·87	795·27	835·51	78·64	73·04
24½	719·68	758·01	798·34	839·97	81·96	75·93
25½	—	759·58	801·24	844·53	84·95	78·99
26½	719·16	760·98	804·18	849·08	88·10	82·18
27½	719·16	762·35	807·35	853·81	91·46	84·96
28½	719·16	764·12	810·53	858·54	94·42	88·12
29½	719·16	765·69	813·78	863·45	97·76	90·93
30½	—	767·61	817·30	868·53	100·92	94·09
31½	719·68	769·36	820·67	873·54	104·18	97·26
32½	720·03	771·27	824·23	878·58	107·31	100·17
33½	720·38	773·37	827·79	883·79	110·42	103·12
34½	720·90	775·46	831·59	889·10	113·64	106·15
35½	721·46	777·64	835·25	894·50	116·86	109·19
36½	722·12	779·91	839·27	900·02	120·11	112·23
37½	722·99	782·27	843·15	905·47	123·20	115·31
38½	723·77	784·71	847·25	911·14	126·43	118·18
39½	724·72	787·29	851·28	916·98	129·69	121·36
40½	725·77	789·78	855·62	922·61	132·83	124·41
41½	726·90	792·57	859·96	928·62	136·05	127·24
42½	728·21	795·37	864·15	934·39	139·02	130·28
43½	729·60	798·34	868·82	940·59	142·25	133·08
44½	730·82	801·31	873·26	946·70	145·39	136·31
45½	732·39	804·28	878·01	953·04	148·76	139·27
46½	733·96	807·43	882·56	959·20	151·77	142·29
47½	735·53	810·75	887·57	965·54	154·79	145·13
48½	737·27	814·07	892·25	972·06	157·99	148·14
49½	739·10	817·40	897·43	978·58	161·18	151·17
50½	741·11	820·89	902·20	985·20	164·31	154·18
51½	743·02	824·40	907·67	992·07	167·67	157·05
52½	745·20	828·15	912·65	998·86	170·71	160·28
53½	747·38	831·75	918·21	42005·74	173·95	163·09
54½	749·64	835·77	923·48	012·63	176·86	166·12
55½	751·92	839·62	928·93	019·69	180·07	169·02
56½	754·53	843·61	934·59	026·93	183·32	171·96
57½	—	847·73	940·28	034·00	186·27	174·96
58½	—	851·97	945·91	041·42	189·45	177·76
59½	—	856·26	951·98	—	—	180·78
60½	—	860·64	—	—	—	—

Table 4 (*cont.*)

$^2\Sigma \rightarrow {}^2\Pi_{\frac{1}{2}}$

J''	P_1	Q_1	R_1	$^SR_{21}$	$\Delta_2 T_1'(J)$	$\Delta_2 T_1''(J)$
2½	—	—	—	41981·58	—	—
3½	41965·01	—	41976·25	984·22	11·24	—
4½	962·99	—	977·44	987·05	14·45	15·25
5½	961·00	—	978·76	989·96	17·76	18·24
6½	959·20	—	980·18	992·95	20·97	21·27
7½	957·49	41968·71	981·58	996·11	24·09	24·31
8½	955·86	968·71	983·18	999·23	27·32	27·31
9½	954·27	968·71	984·83	42002·57	30·56	30·32
10½	952·86	—	986·59	005·74	33·73	33·37
11½	951·46	—	988·45	009·45	36·99	36·37
12½	950·22	—	990·39	012·98	40·17	39·46
13½	948·99	—	992·33	016·51	43·34	42·45
14½	947·94	970·40	994·54	020·31	46·60	45·45
15½	946·98	971·00	996·74	024·06	49·86	48·54
16½	946·00	971·65	999·03	027·81	53·03	51·52
17½	945·22	972·49	42001·34	031·79	56·12	54·51
18½	944·52	973·25	003·85	036·02	59·33	57·63
19½	943·81	974·30	006·46	040·11	62·65	60·53
20½	943·27	975·31	009·08	044·51	65·81	63·63
21½	942·83	976·47	011·82	048·76	68·99	66·60
22½	942·48	977·70	014·74	053·09	72·26	—
23½	—	979·01	017·57	057·69	—	—
24½	—	980·42	020·65	062·37	—	75·70
25½	941·87	981·95	023·75	067·13	81·88	78·78
26½	941·87	983·51	026·93	071·85	85·06	—
27½	—	985·20	030·20	076·81	—	—
28½	—	987·05	033·57	081·61	—	87·72
29½	942·48	988·85	030·10	086·73	94·62	90·74
30½	942·83	990·92	040·64	091·87	97·81	93·83
31½	943·27	992·84	044·25	097·19	100·98	96·83
32½	943·81	995·10	048·05	102·50	104·34	99·73
33½	944·52	997·28	051·85	107·95	107·33	102·83
34½	945·22	999·60	055·75	113·50	110·53	105·85
35½	946·00	42001·94	059·73	119·10	113·73	108·77
36½	946·98	004·51	063·91	124·68	116·93	111·79
37½	947·94	007·05	068·13	130·53	120·19	114·92
38½	948·99	009·78	072·41	136·48	123·22	117·81
39½	950·32	012·63	076·81	142·34	126·49	120·58
40½	951·63	015·45	081·24	148·43	129·61	123·77
41½	953·04	018·49	085·84	154·60	132·80	126·79
42½	954·45	021·43	090·37	160·65	135·92	129·81
43½	956·03	024·62	095·23	—	139·20	132·75
44½	957·72	027·86	100·02	—	142·30	135·59
45½	959·44	031·25	104·98	—	145·34	138·62
46½	961·40	034·60	109·95	—	148·55	141·67
47½	963·31	038·17	115·05	—	151·74	144·58
48½	965·37	041·80	120·20	—	154·83	147·57
49½	967·48	045·47	125·56	—	158·08	150·60
50½	969·60	049·36	130·89	—	161·29	153·50
51½	972·06	053·25	136·39	—	164·33	156·53
52½	974·36	057·14	141·90	—	167·54	159·47
53½	976·92	061·21	147·58	—	170·66	162·44
54½	979·46	065·45	153·31	—	173·85	165·38
55½	982·30	069·70	—	—	—	168·38
56½	984·93	074·33	—	—	—	—
57½	987·93	078·76	—	—	—	—
58½	990·92	083·19	—	—	—	—
59½	994·01	087·97	—	—	—	—
60½	997·10	092·58	—	—	—	—

621.317.755

TIME-MARKING A CATHODE-RAY OSCILLOGRAPH BY HARMONICS

By L. F. RICHARDSON, M.A., D.Sc., F.Inst.P., F.R.S.

Received September 6, 1934. *Read in title December* 7, 1934.

ABSTRACT. A docile time-marker was arranged in which the oscillograph acted also a a triode giving retroaction in a tuned circuit where one of the harmonics of the trace frequency was selected and applied to the Wehnelt cylinder so as to modulate the focus o the trace.

---

§ 1. GENERAL DESCRIPTION

A former paper(1) described time-marks made by periodically unfocusing the electron beam by applying an oscillating potential to the cylindrical shield which surrounds the filament of the oscillograph. The time-marking potential was produced by a triode oscillator of an ordinary kind. It subsequently occurred to me that the oscillograph might itself serve as a triode. The circuit shown in figure 1 proved to be extremely convenient, after various difficulties had been overcome in the way that will be described.

The Wehnelt shield *G* of the oscillograph behaved like the grid of a triode, controlling the anode current, so as to produce retroaction; but of course its potential to the filament is of the opposite sign to that in a triode. No special coupling was introduced between the phenomenon and time-marking circuit. There is however inevitably a small capacitance-coupling inside the oscillograph and at its base.

In the previous method, where a quite independent oscillator was used to interrupt the trace, the ratio of the two frequencies, that of the time-marker and that of the phenomenon, had to be very accurately adjusted in order to hold the time-marks steady enough for photography. A change in the ratio of frequencies changed the speed of the marks and they were apt to race so fast as to become invisible. That is to say, the previous arrangement behaved like a stroboscope.

The present apparatus has several types of actual or possible behaviour. *Type I.* If retroaction could be made sufficient to maintain free oscillation of the circuit *NK*, we should presumably have again the headstrong stroboscopic behaviour of the independent oscillator. *Type II.* Actually retroaction has been just not enough for steady free oscillations, and the behaviour of the apparatus was correspondingly docile. The time-marks never moved too fast to be seen. When the frequency of either the light-spot or the time-marker has been slightly changed, and left so, the marks have not acquired a new speed, they have acquired a new stationary position and sometimes a new distinctness. This steadiness is a great advantage for photography.

The advantage however is not all in favour of a time-marker with a decrement; for the greatest attainable number of time-marks on the trace has been much smaller, only 15; whereas with steady free oscillation 196 were attained. The greatest number of marks obtainable with the circuit of figure 1 appeared to be dependent on the highest harmonic in the voltages applied to the deflector plates; for when these voltages were pure sine waves only one blur could be obtained on the trace. The finest attainable subdivision of time is thus automatically almost sufficient to analyse the finest detail in the phenomenon.

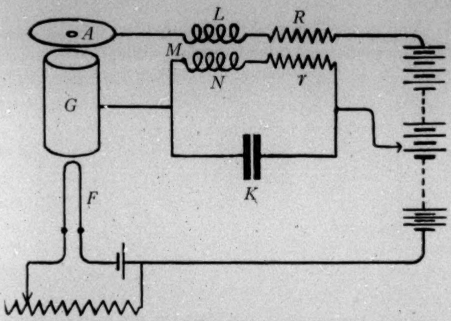


Figure 1.

There are two varieties of type II depending on whether there is a jerk at one part of the trace or not. A jerk in a periodic phenomenon can of course be analysed by Fourier methods into harmonics. Nevertheless there is something in the relations between the phases and amplitudes of the successive harmonics which distinguishes a jerk from a ripple.

Variety IIA. The free oscillation of the time-marker was damped, but it was started once in every period of the phenomenon by a jerk, and decreased during the rest of the period. This is shown in plate I, figure 1 which consists of three oscillograms superposed, the time-marker frequency being 11 times that of the trace for one oscillogram, $11\frac{1}{2}$ for the next, and 12 times for the last. The light-spot travelled in the sense $\alpha\beta\gamma\delta\epsilon$ completing the circuit 1821 times per sec. The exposure was about 15 sec. The thinness of the trace between α and β shows that the spot was moving at its fastest there. A time-mark, a blur, stands near γ; and however the capacitance K was varied there always was a time-mark near γ. As the capacitance K was increased the marks spread along the trace $\gamma\delta\epsilon$, the mark near γ remaining almost stationary, the others being more displaced as they were farther along the trace. The conspicuousness of the marks decreased from δ to ϵ and from ϵ to α. This indicates that the rapid portion $\alpha\beta$ acted like an electrical jerk in starting the oscillation in the circuit NK; and that the circuit NK was oscillating with a decrement.

On reversal of the sign of the mutual inductance M by turning over of one coil, the marks became much feebler on $\delta\epsilon$ though still plain on $\gamma\delta$. This showed that the oscillograph had indeed been giving retroaction, which had tended to maintain the oscillations started by the jerk. As the capacitance K in the time-marker was continuously varied, the blurs not only moved along the trace but periodically waxed and waned in distinctness. The values of K producing maxima of distinctness were found to be those for which the time-marker-frequency was an integral multiple of the trace-frequency.

Variety II B. The free oscillation of the time-marker was damped; the oscillation applied to the deflector plates contained harmonics, but no conspicuous jerk. Figure 2 of the plate shows, superposed, three oscillograms in which the light-spot travelled in the clockwise sense 9800 times per sec. The frequency of the time-marker was severally 5, $5\frac{1}{2}$ and 6 times that of the trace.

When the frequency of the time-mark has been changed, all the marks have taken up new positions. Unlike variety II A, variety II B gave no stationary mark. Certain tuning-capacitances K gave marks which were specially conspicuous, and these values of K corresponded to harmonics of the trace-frequency. This was shown by plotting G. W. O. Howe's well-known diagram(6) of K against n^{-2} where n is the number of blurs on the trace. The points lay nearly on a straight line.

The frequency of the time-marks was thus obtained as a whole multiple of the known frequency of the trace. In this way the condenser K was calibrated for frequency. On account of self-capacitance, and of the coupling M/LN, the values of K and N determined at low frequencies had to be regarded as giving merely a rough check on the time-marking frequency.

To produce satisfactory marks, the shield voltage was first adjusted by battery so that the trace was slightly blurred. When the oscillatory voltage of the circuit KN was superposed some parts of the trace went into focus, other parts became more blurred. The amplitude of the time-marking voltage was estimated by altering the superposed steady voltage of the focusing battery until the time-marks alternately appeared and disappeared; and was thus found to be about 1 V. As this is a small fraction of the range of good focus, it is certain that there was one blur, not two, in each complete oscillation of the time-marking voltage.

To avoid any appreciable deflection of the cathode-rays by the stray field from the time-marking coils it was found to be quite sufficient to remove the coils to a distance of 1 m. from the tube and to orient them so that the electrons moved along the lines of magnetic force, for the central ray. The particulars thus far given may suffice to justify the results obtained by this method and set out in another memoir.

§ 2. FURTHER PARTICULARS

This section is appended for those who wish to use the method in the laboratory.

Coils. Some guidance as to the kind of coils required for L and N was provided by the well-known theory(2) of the maintenance of oscillations in the similar circuit having a triode instead of an oscillograph. For maintenance or growth,

$$M \geqslant \frac{rK}{g} \text{ approximately} \qquad \text{......(1)},$$

where M is the mutual inductance and r is the resistance of the condenser plus that of the coil N both at the resonant frequency and $g = \partial u_a / \partial e_g$ in which u_a is the anode

PHYSICAL SOCIETY, VOL. 47, PT. II (RICHARDSON)

Figure 1.

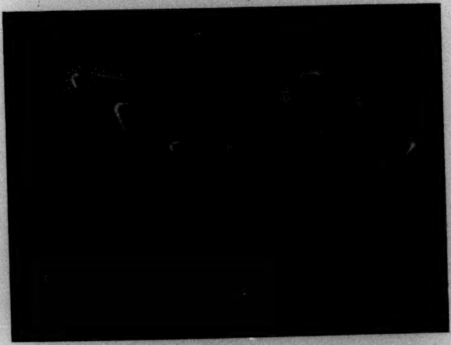


Figure 2.

[*To face p.* 260

current and e_g the potential between grid and filament. In order to increase g it was found beneficial to increase the filament current to 0·93 A. and to decrease the anode voltage e_a to 265 V. This done, g was found to be 7×10^{-5} ampere volt^{-1} for an oscillograph supplied by Standard Telephones and Cables, Ltd. and marked type 4018-A.

Thus g for the oscillograph was of the order of $\frac{1}{30}$ of the g customary in small triodes. Consequently the coils had to have small decrements if the oscillations were to be maintained.

We have by the theory of the tuned circuit at frequency ν,

$$K = \frac{1}{(2\pi\nu)^2 N} \text{ approximately} \qquad \text{......(2).}$$

On elimination of K between (1) and (2),

$$M \geqslant \frac{r}{(2\pi\nu)^2 Ng} \qquad \text{......(3).}$$

If M is unknown we have anyway the restriction

$$\sqrt{(LN)} > M \qquad \text{......(4).}$$

Eliminating M between (3) and (4),

$$\frac{L^{\frac{1}{2}} N^{\frac{3}{2}}}{r} \geqslant \frac{1}{g\,(2\pi\nu)^2} \qquad \text{......(5).}$$

On substitution of numerical values it was seen that this condition was not easy to satisfy. Suitable coils were either costly to buy or troublesome to make; but unlike triodes and batteries they should last a lifetime.

Several coils were tried for L and N. The frequency required was of the order of 10^4 c./sec. or 10^5 c./sec. Secondaries of old induction coils or slices thereof were useless; they had too low an open-circuit frequency, about 2×10^3 c./sec. and 3×10^3 c./sec. Radio coils, sold for broadcast reception at 2×10^5 c./sec., had too small a time-constant N/r. Best wave-meter coils should do; but actually some coils were made for the purpose, after studying information about self-capacitance and high-frequency resistance given in writings by G. W. O. Howe$^{(5)}$, D. W. Dye$^{(4)}$ and E. B. Moullin$^{(7)}$.

Three particular coils formed serviceable combinations when arranged as explained in the table.

Reference mark	Self inductance at 50 c./sec. (henries)	Open-circuit frequency (c./sec.)	Coil resistance at 10° C.		Resistance r at 9800 c./sec. (Ω.)
			at 0 c./sec. (Ω.)	at 9800 c./sec. (Ω.)	
A = R 33/2	0·124	$4{\cdot}5 \times 10^4$	102	130 ± 5	< 170
B = R 33/3	0·00857	$1{\cdot}5 \times 10^5$	2·4	2·9	3·4
C = Igranic Honeycomb	0·0100	4×10^5 Makers' value	23·6		

The inductances were measured by the aid of a Campbell variable mutual inductance. The high-frequency resistances were measured both by distuning and by the added resistance method; as described by Moullin$^{(7)}$. The dielectric in the condensers was mica or air.

Combination for time-marker frequencies between 10^4 *and* 3×10^4 *c./sec.* Tuning-coil A, anode-coil B. At the lowest frequency in this range

$$\frac{r}{(2\pi\nu)^2\, Ng} < 0\cdot0052 \text{ H.},$$

whereas $M=0\cdot0095_5$ H., so that the condition for growth was satisfied at 9800 c./sec., and probably throughout the range of frequency.

Combination for time-marker frequencies between 2×10^4 *and* 10^5 *c./sec.* Tuning-coil B, anode-coil C. At the lowest frequency in this range, r being estimated as less than 6 ohms,

$$\frac{r}{(2\pi\nu)^2\, Ng} < 0\cdot000{,}65 \text{ H.},$$

whereas $M=0\cdot001_4$ H. or more. Again the condition for growth was satisfied.

Yet actually, for both combinations, the time-marker oscillated with a decrement. So there must be in the oscillograph some dissipation of energy not occurring in a triode. The high resistance of the return path from the fluorescent screen is conceivably the cause of the dissipation.

Order of connections. A minor difficulty was that when the anode connections were altered while the filament was lit, the anode fuse melted, at 0·1 A. The energy presumably came from the magnetic field of the coil L. It was necessary to make the connections in the order opposite to that recommended by the makers⁽³⁾, putting on the anode volts permanently while the filament was cold and then slowly increasing the filament current. The reverse procedure was adopted for disconnecting.

Focussing battery. For adjusting the battery-voltage on the Wehnelt shield, steps of 1·5 V. were rather large, but 0·5 V. steps were small enough. The latter were provided by placing small accumulators in opposition to dry Leclanché cells.

Intermittent vision. Faint time-marks, which might pass unnoticed if persisting steadily, became clearly visible when the time-marker capacitance K was switched repeatedly in and out of resonance.

§ 3. ACKNOWLEDGMENT

This paper is a bye-product of a research for which apparatus was provided by the Government Grant Committee. This loan is gratefully acknowledged.

REFERENCES

(1) L. F. Richardson. *Proc. phys. Soc.* **45**, 135–41 (1933).
(2) L. B. Turner. *Wireless*, pp. 265, 266 (Camb. Univ. Press, 1931).
(3) Standard Telephones and Cables, Ltd. Bulletin G 577/3 "Improved Cathode Ray Oscillograph," Type 4018-A.
(4) D. W. Dye. *Dict. App. Physics*, **2**, "Radio Frequency Measurements."
(5) G. W. O. Howe. *J. Instn elect. Engrs*, **60**, 67–72 (1921).
(6) G. W. O. Howe. *Proc. phys. Soc.* (1912) quoted in E. B. Moullin's *Radio Frequency Measurements*, p. 338 (Griffin, 1931).
(7) E. B. Moullin. *Radio Frequency Measurements* (Griffin, 2nd ed., 1931).

551.594.11

IONIZATION CHARTS OF THE UPPER ATMOSPHERE, PART II

By G. MILLINGTON, Marconi's Wireless Telegraph Co., Ltd.

Communicated by Prof. S. Chapman, August 30, 1934. *Read December* 21, 1934.

ABSTRACT. This paper discusses an envelope correction to some ionization charts previously published, and in addition presents the new charts on a circular projection to help in the study of conditions in the polar regions.

---

§ 1. INTRODUCTION

This paper is in the nature of an appendix to a previous paper, here referred to as "part I," which was published some time ago(1), and is mainly concerned with a small correction to the charts contained therein. It was pointed out that in order to reduce the amount of work to reasonable dimensions the simplifying assumption was made that the density, time (ν, ϕ) curve for the height z_0 at which the noon rate of ionization was a maximum could be taken as the envelope of the system of (ν, ϕ) curves for varying values of the height z. It was realized that the error involved would be greatest in the sunrise region, but it was thought that for the values of the σ_0 chosen, namely 0·5 and 1, the error would not be serious. Detailed comparison of the theoretical charts with experimental data has shown, however, that there is a consistent discrepancy in this critical region, the charts always yielding density-values which are too low. The error introduced by the assumption may thus be greater than was anticipated, and it is important to know whether it is responsible for the whole of the discrepancy or whether there is a residual effect to be accounted for in some other way.

ν, ϕ, z_0

z

It was therefore decided to investigate the problem at least for one set of conditions, e.g. for latitude 60° in winter, where the effect would be most noticeable. Accordingly a number of (ν, ϕ) curves were worked out for different values of z. Now in figure 3 of part I the rate-of-ionization function $F(z, \chi)$ is plotted for a number of values of z, where χ is the zenithal angle of the sun.

F

χ

Actually z is there called z_0 and is expressed in terms of χ_0 by the relation $z_0 = \log_e f(R, \chi_0)$, the curves being drawn for a series of values of χ_0, the noon zenithal angle of the sun. The curves are only drawn for values of χ beyond their respective χ_0 values, and it will be seen that each curve lies entirely outside all the curves for smaller values of χ_0. If, however, the curves are continued backwards it will be found that each curve crosses all the curves before it. For the chosen latitude and season the minimum value of χ is the appropriate value of χ_0, and beyond χ_0 the F curves for values of z smaller than z_0 will all lie inside the z_0

R, χ_0

curve and need not be considered in determining the envelope of the (ν, ϕ) curves; for it is obvious that if the F curves for two values of z do not intersect then the corresponding (ν, ϕ) curves will not intersect either, and we can ignore the inner curve.

Fortunately a little work soon suggested an easy method of making a close approximation to the required envelope, by means of which it was possible to work out the complete modified charts with less labour than the original charts required. The fact already mentioned, that the curves of figure 3 in part I if continued back intersect, suggests that the set of curves has an envelope which passes through the initial points of each of the curves. From equation (1) of part I we have for the form of the function F

$$F = \exp[1 - z - \exp(-z).f(R, \chi)],$$

and the equation $\partial F/\partial z = 0$ for eliminating z to obtain the equation of the envelope gives

$$-1 + \exp(-z).f(R, \chi) = 0,$$

and on substitution in F for the equation of the envelope,

$$F = \exp(-z) = \frac{1}{f(R, \chi)}.$$

From equation (7) of part I it follows at once that when $\chi = \chi_0$,

$$F(z_0 \chi) = 1/f(R, \chi_0),$$

so that the envelope passes through the initial points of the curves of figure 3 of part I, as was anticipated. Now since the function $f(R, \chi)$ increases upwards from unity with increasing values of χ, the solution for the envelope is only valid for positive values of z (it being remembered that z is not the actual height h but is referred to a height h_0 as datum level). This corresponds to the easily verified fact that as z increases negatively the F curves lie wholly inside one another, but as they are all inside the curve for $z = 0$ they can be neglected as far as our problem is concerned.

Now if the fundamental differential equation, given in (4) of part I,

$$\sigma_0 \frac{d\nu}{d\phi} + \nu^2 = F,$$

is solved not for a given height z but for the envelope function F or $1/f(R, \chi)$, then the (ν, ϕ) curve obtained must lie outside all the possible (ν, ϕ) curves for various values of z. Calculation shows moreover that if it is not the true envelope it is at any rate a very close approximation for widely different conditions of season and latitude, especially in the critical region where the correction is of most importance. Although the mathematical process involved may have no physical justification, it can be seen that the method suggests what in a general way must be true, that as time progresses the position where the maximum density occurs will tend to follow the variation in height at which the rate of ionization is a maximum. Moreover the method suggests a possible interaction between neighbouring regions which may occur as a process of diffusion of ions, and we are probably not justified in seeking to approximate closer to the envelope of the (ν, ϕ) curves. The modified charts have therefore been worked out on this basis, and thus the labour required is actually reduced owing to the use of a single F curve instead of having to interpolate from a set of $F(z_0\chi)$ curves. In any case the charts now set an upper limit to the theoretical ν values.

§ 2. CALCULATION OF THE CORRECTED CHARTS

Much of the material already obtained and outlined in § 5 of part I can be used, but a graph of the envelope function F or $1/f(R, \chi)$ is needed and is given in figure 1. In the calculations for solving the differential equation, F can be neglected when it is less than 0·001, i.e. when χ is greater than about 102°, but the region between $\chi = 90°$ and 102° is very important, especially in high latitudes in winter when χ_0 itself is large. Now as the curve for F does not take account of the fact that as χ increases above 90° the curvature of the earth cuts off the sun's radiation from the lower regions of the atmosphere, we must verify that in practice the F function does exist at least as far as $\chi = 102°$. If we write $\chi = 90 + \epsilon$ then at a height h the sun's radiation is cut off when ϵ exceeds the limiting value $\epsilon_{\text{max.}}$ given by $\sec \epsilon_{\text{max.}} = 1 + h/R$*, and in figure 2 $\chi_{\text{max.}}$ is plotted as a function of h. Although the curve soon settles down to very small increases of $\chi_{\text{max.}}$ for large increases of h, χ can be greater than 102° provided that h is greater than 145 km. Since in the construction of the charts we are mainly concerned with the upper or F region of the Heaviside layer, h is actually greater than this value, so that in practice the increased attenuation of the sun's radiation due to the longer path through the earth's atmosphere reduces the F function to a negligible value before the radiation is actually cut off by the curve of the earth.

The solution of the differential equation follows along the lines given in part I and the resulting set of (ν, ϕ) curves is given in figures 3 to 5 where for convenience ϕ has been converted from radians to Greenwich mean time. They are only given for the value $\sigma_0 = 0{\cdot}5$, as experience has shown that this value corresponds with present conditions rather than $\sigma_0 = 1$ as suggested in part I, while it may be useful later to construct charts for $\sigma_0 = 0{\cdot}25$. From these curves ν was plotted as a function of the latitude for a series of values of ϕ as this was found useful for interpolating to obtain the (ν, ϕ) curve for any other latitude and was an aid towards the accurate construction of the charts from the curves. The new charts for $\sigma_0 = 0{\cdot}5$ are given in figures 6 and 7.

Comparison between the new and old charts shows that while the positions of the contour lines of equal density are scarcely changed round mid-day, the region after sunrise where the lines are crowded together and represent the rapid early-morning increase of density is moved noticeably towards the sunrise line. But although in other parts the charts agree well with experimental data there is still some discrepancy in the sunrise region sufficient to suggest that there is an effect not accounted for by the theory.

* R here is of course not Chapman's parameter in $f(R, \chi)$ but the radius of the earth.

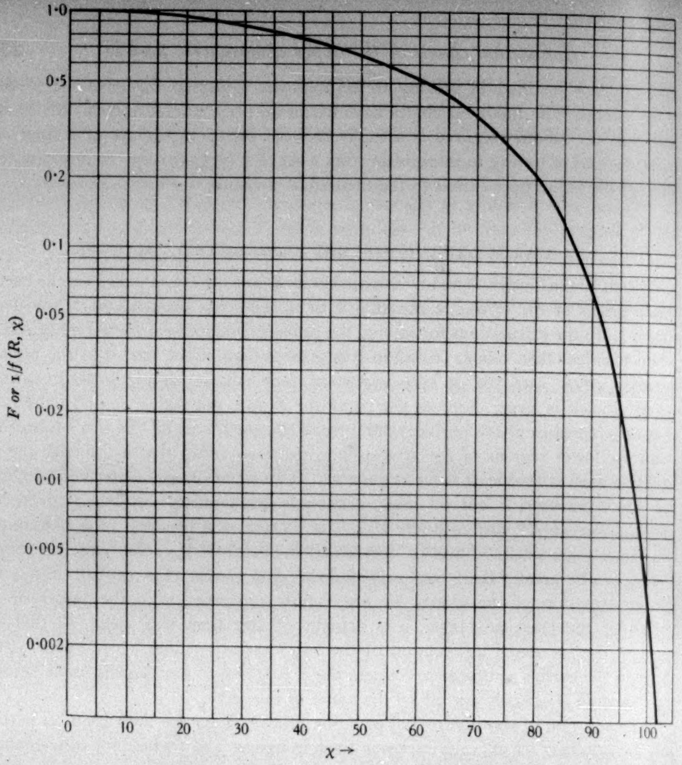


Figure 1. The envelope function F or $1/f(R, \chi)$.

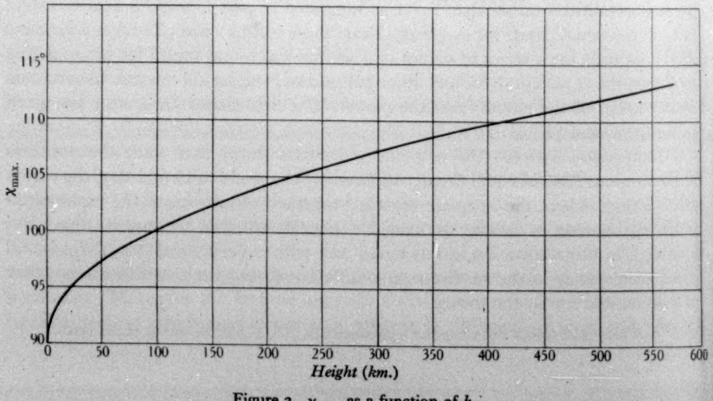


Figure 2. $\chi_{\text{max.}}$ as a function of h.

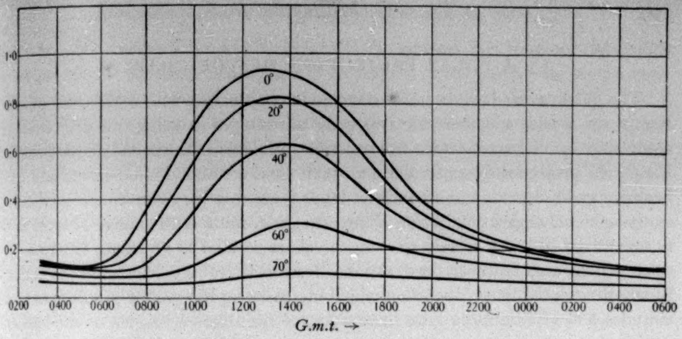


Figure 3. Winter, $\sigma_0 = 0{\cdot}5$. Latitudes 0, 20, 40, 60, 70°.

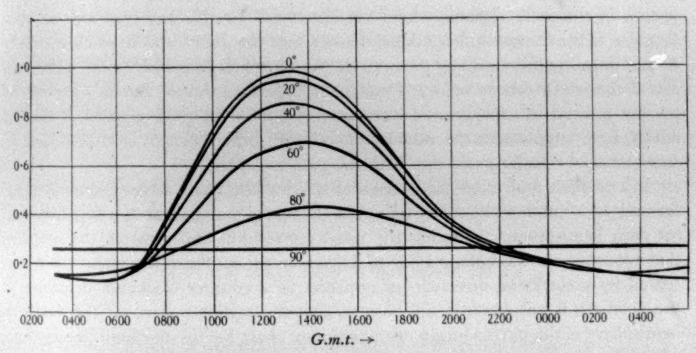


Figure 4. Equinox, $\sigma_0 = 0{\cdot}5$. Latitudes 0, 20, 40, 60, 80, 90°.

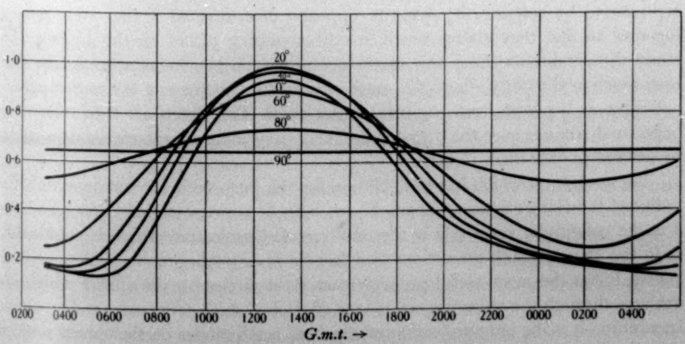


Figure 5. Summer, $\sigma_0 = 0{\cdot}5$. Latitudes 0, 20, 40, 60, 80, 90°.

§ 3. A POLAR PROJECTION OF THE CHARTS

The charts so far have been constructed in conjunction with a Mercator projection of the world, and there are two great advantages in using this projection to study short-wave long-distance transmission problems with the aid of such charts. Firstly the projection is conformal, i.e. small areas are shown undistorted and true bearings are indicated, and secondly it lends itself to a very convenient method of studying the change in ionization along any given route with change of local time at some fixed place by merely moving the chart parallel to its length from east to west. There are, however, two serious drawbacks to the Mercator projection. Firstly there is the linear distortion which increases rapidly as one proceeds to high latitudes and gives quite a false impression of the relative lengths in the various grades of density of the chart for any route running more or less from north to south. Arising from this we have secondly the complete break-down of the projection in the polar regions, where the charts fail to give any true idea of what happens to the contour lines of equal density over the polar cap. In order therefore to study the conditions along great-circle routes which pass near to the poles, e.g. round-the-world echoes on the London-to-Capetown route, and in view of the present interest in experimental work carried out in the polar regions, it seemed advisable to supplement the existing charts with others constructed according to some form of circular projection with the pole as centre.

In any such projection the process of moving the chart across the map transforms into a rotation about the pole, but obviously we shall need a separate map for each hemisphere. Adopting the usual convention and drawing the maps so that a journey eastward along a line of latitude in the southern hemisphere is represented by a clockwise direction, as opposed to a counter clockwise direction in the northern hemisphere, it will be seen that the equinox chart for the southern hemisphere is the mirror image of the equinox chart for the northern hemisphere, while the southern chart for the December solstice is the mirror image of the northern chart for the June solstice and vice versa. The rotation of the earth is represented by rotating the maps in opposite directions as if they were geared together so that they always touch at corresponding places on the equator. To study the conditions along any north-to-south route the maps are fixed so that they touch at the point where the route crosses the equator and the route appears unbroken as it passes from one map to the other. The charts are then rotated in opposite directions over the maps, and if necessary some simple mechanism could be devised to gear them together.

The projection which has been chosen for the purpose is the one in which the radius of the circle representing any line of latitude is proportional to tan $\frac{1}{2}\theta$ where θ is the colatitude, since this is the only circular projection which is conformal. It is the stereographic projection used largely in crystallography, and as is well known it has the most useful property that great circles on the sphere which do not pass through the poles project as arcs of circles which terminate on diameters. In addition it is the only projection in which all small circles on the sphere project as circles. (Of course if the small circle cuts the equator this property only holds if we project the southern hemisphere with respect to the north pole on the same map as the northern hemisphere or vice versa.) This property is useful if we wish to study the ionization at places in relation to their distances from some fixed point other than the pole, e.g. in discussing effects in the neighbourhood of the magnetic pole or the skip-distance phenomena observed at any particular place. The linear distortion with this projection is never more than 2 to 1 and owing to the conformal property is independent of direction. Actually, when referred to the equator as standard, distances in colatitude θ must be multiplied by $1+\cos\theta$, so that it is fairly easy to estimate the true lengths of any route which lie in the various grades of the chart. In practice the size of the map is chosen so that the polar regions where areas are reduced fourfold are represented on a reasonably large scale.

As the projection is well known it is not thought necessary to give here any detailed methods of drawing in circles on it, but the following suggestions may be helpful. If we wish to draw the projection of a small circle of radius δ degrees δ
round the point whose longitude is ϕ and colatitude is θ, the points $(\phi, \theta-\delta)$ and $(\phi, \theta+\delta)$ which can be put on the projection at once by the tan $\frac{1}{2}\theta$ rule will define a diameter of the required circle which can then be constructed from geometrical principles. This holds even if the small circle cuts the equator, but in this case we have in addition the fact that where it cuts the longitude is $(\phi\pm\psi)$ where $\cos\psi=\cos\delta/\sin\theta$. If we wish to draw the great circle joining two points whose ψ
longitudes are ϕ_1 and ϕ_2 respectively and whose colatitudes both measured from the north pole are θ_1 and θ_2 respectively, then the great circle cuts the equator at longitudes ϕ and $(\phi+\pi)$ where $\tan\phi=a_2/a_1$

and $$a_1=\sin\theta_1\cos\theta_2\cos\phi_1-\sin\theta_2\cos\theta_1\cos\phi_2,$$ a_1

while $$a_2=\sin\theta_1\cos\theta_2\sin\phi_1-\sin\theta_2\cos\theta_1\sin\phi_2$$ a_2

and the nearest points of approach to the poles will be given by $(\phi\pm\frac{1}{2}\pi, \theta_{\text{min.}})$,

where $$\tan\theta_{\text{min.}}=a^{-1}\sin\theta_1\sin\theta_2\sin(\phi_2-\phi_1)$$

and $$a^2=a_1{}^2+a_2{}^2.$$ a

From these points the circle can be constructed.

Lastly it can be shown that the great circle through any point (θ, ϕ) which makes an angle δ with the great circle passing through the point and the poles cuts the equator at longitude $(\phi\pm\psi)$ where $\tan\psi=\cos\theta\tan\delta$, the sign depending upon ψ
the direction in which δ is taken.

The charts for the northern hemisphere for $\sigma_0=0{\cdot}5$ are shown in figures 8 to 10 and are based on the corrected charts of figures 6 and 7 with extra contour lines for $\nu=0{\cdot}13$, $0{\cdot}05$ and $0{\cdot}02$ to give more detail in the polar regions in winter. In figures 11 and 12 are given outline maps of the two hemispheres for use with the charts. The charts show well how little seasonal variation there is on the equator. In summer the minimum density over the hemisphere occurs on the equator about twenty minutes before sunrise and corresponds to $\nu=0{\cdot}13$. At the

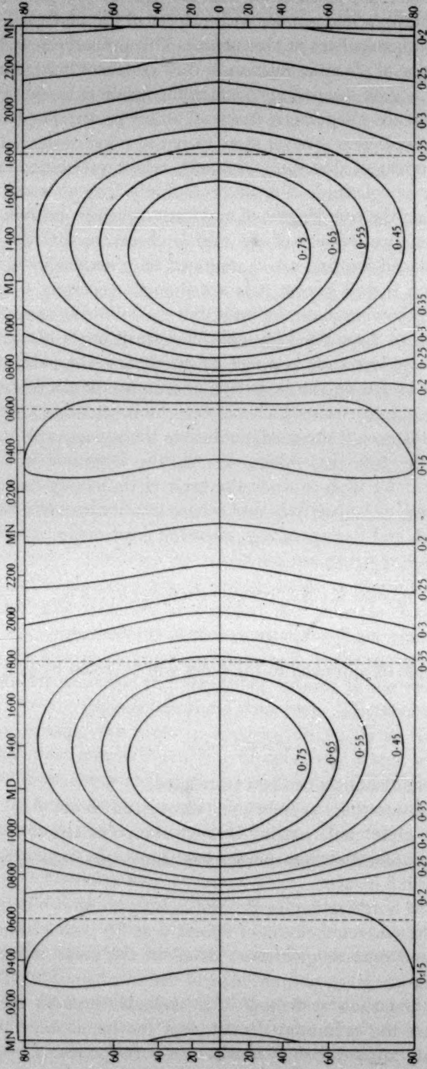


Figure 6. Equinox chart, $\sigma_0 = 0{\cdot}5$.

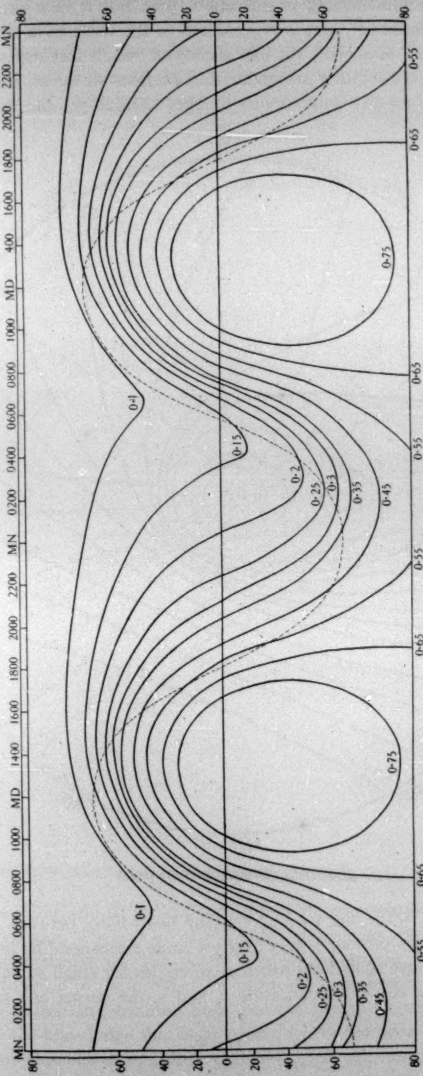


Figure 7. Winter chart, $\sigma_0 = 0{\cdot}5$.

equinox the minimum density is still about 0·13 but it now extends from the equator up to latitude 60°, and is represented in the chart by a single thick line. In winter this line has opened out into a contour which just reaches the equator (corresponding to the summer minimum) and encloses all the polar regions within latitude 68°. The $\nu = 0{\cdot}10$ contour just touches latitude 70° but does not extend below latitude 40°, while the $\nu = 0{\cdot}05$ contour lies wholly between latitudes 70° and 80°. The $\nu = 0{\cdot}02$ contour is very nearly a circle corresponding to latitude 80°. That the contour lines should approximate to circles for small values of ν can be seen by writing the equation connecting ν_s and ν_r the sunset and sunrise values in the form

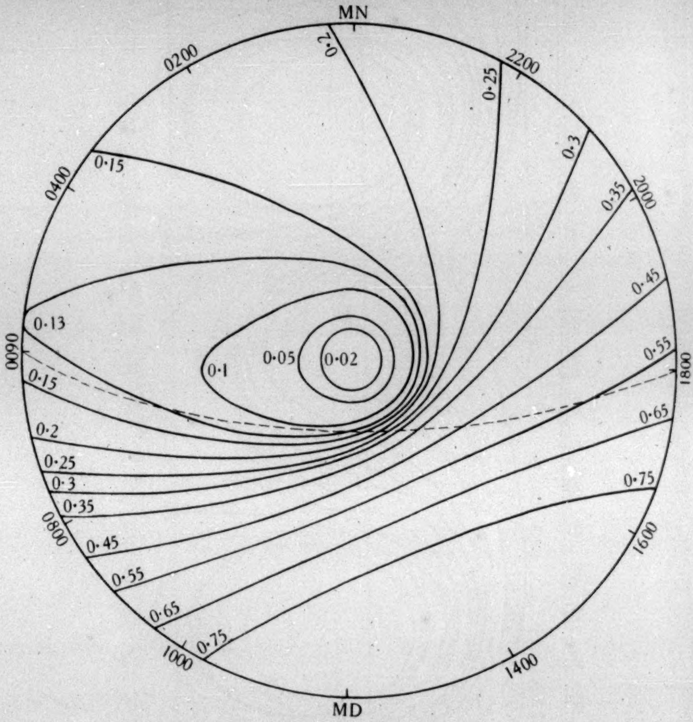


Figure 8. Winter chart, $\sigma_0 = 0{\cdot}5$.

$$\frac{\nu_s}{\nu_r} = 1 + \frac{\phi_r - \phi_s}{\sigma_0} . \nu_s,$$

so that even in the long winter night when $\phi_r - \phi_s$ approaches 2π, ν_s/ν_r differs from unity by less than 0·25 when $\nu = 0{\cdot}02$.

The extra contour lines show how rapidly the maximum density decreases in the region of the pole at mid-winter. At the pole itself the ionization is negligible and it is a matter of some importance to know the extent of the polar cap over which there is effectively no ionization and how the area diminishes on either side of the winter solstice. There is no easy way of determining what are the maximum and minimum densities for any given latitude and season, but we can get an approximate solution to the problem by using the fact already mentioned that for small density values the contour lines become nearly circular about the pole, together with the knowledge that at any place where the noon zenithal angle is χ_0 the maximum value of ν must be less than $\sqrt{\{1/f(R, \chi_0)\}}$ except at the pole where

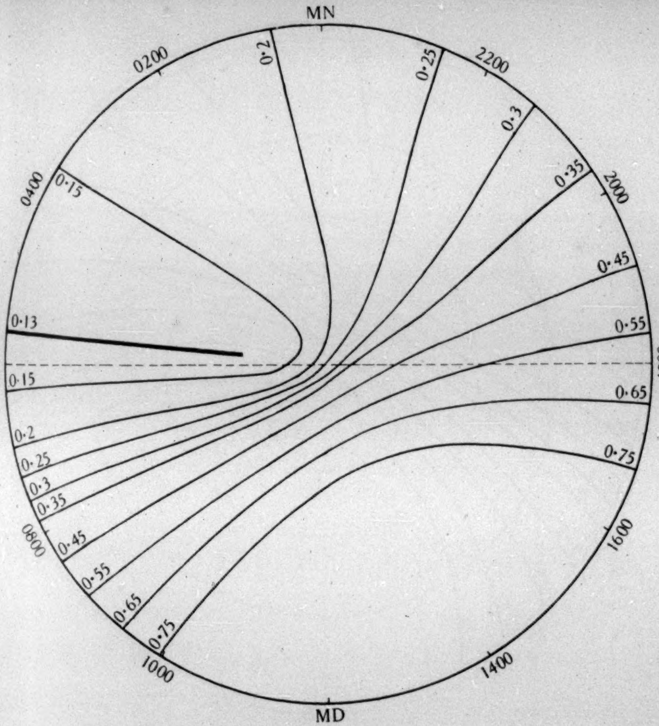


Figure 9. Equinox chart, $\sigma_0 = 0{\cdot}5$.

it is equal to this quantity. Knowing that for a latitude l when the declination of the sun is δ the value of χ_0 is $|\,l-\delta\,|$, we can find at once the upper limit of ν. Thus, as we have seen, ν is about 0·02 at latitude 80° in mid-winter when $\delta=-23{\cdot}5°$, and extrapolation from a graph of $\sqrt{\{1/f(R,\chi)\}}$ shows that ν is probably less than 0·01 at latitude 81·5°, less than 0·005 at 83°, and possibly less than 0·001 at

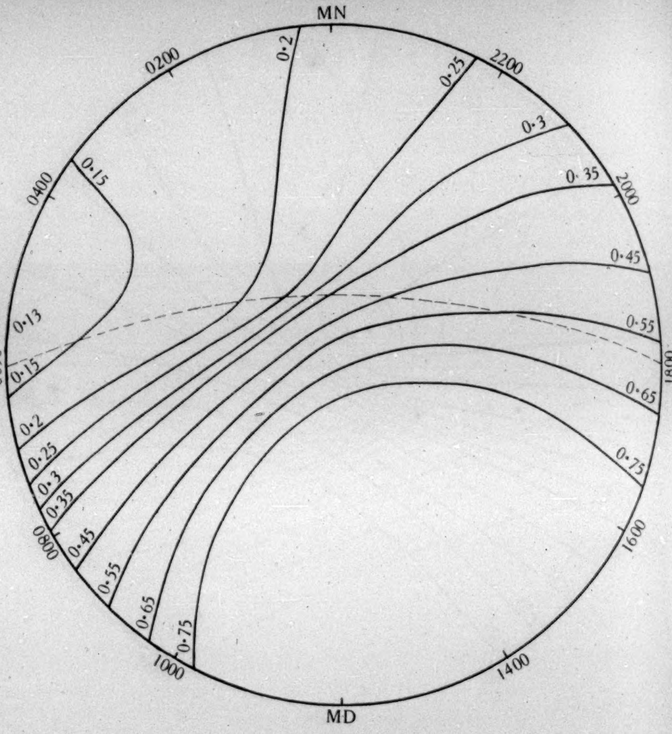


Figure 10. Summer chart, $\sigma_0=0{\cdot}5$.

latitude 86°. In considering the effect of the change of season we have to remember that as we move away from mid winter the minimum density does not remain at the pole. The equinox chart shows the way in which the contour lines spread out across the pole until nowhere over the earth's surface does the density fall below $\nu=0{\cdot}13$. But at first the major effect is the shrinking of the contours of small density into the pole with the contraction and final disappearance of the area of negligible ionization. Now the relation $\chi_0=|\,l-\delta\,|$ gives an approximate rule that as δ changes from the mid-winter value of $-23{\cdot}5°$ the contours for $\nu<0{\cdot}02$ close in to a latitude nearer the pole by the amount of change in δ. Thus when $\delta = -13{\cdot}5^\circ$ the density has everywhere risen above $\nu = 0{\cdot}02$, and this corresponds to a period of about eight weeks on either side of the winter solstice. The density is probably everywhere above $\nu = 0{\cdot}05$ when δ is about -10° since ν becomes $0{\cdot}05$ at the pole when $\delta = -10{\cdot}5^\circ$. In this way the extent and duration of this polar dark area, as we may call it, can be fairly well defined.

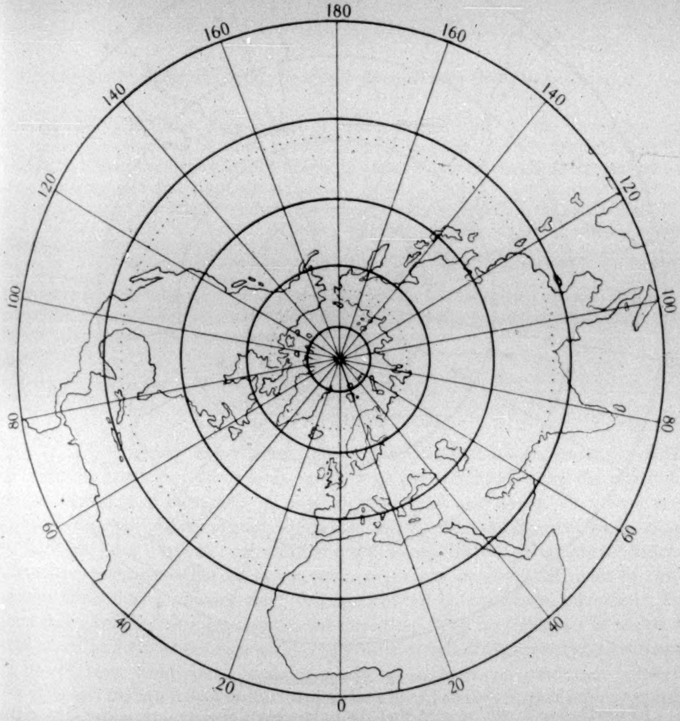


Figure 11. Northern hemisphere.

The significance of this dark area is that it has been worked out on the assumption of ideal theoretical conditions in which it is assumed that there will be no tendency of outside ions to diffuse in towards the pole, and that it implies that at mid-winter short waves could not be transmitted at all on any route crossing this polar area. The theory takes no account of the existence of the earth's magnetic poles and of the large effects which are thereby produced, especially in the northern latitudes into which the electrons from the sun tend to be deflected. Unfortunately the difficulties in the way of making direct observations across the polar cap, especially in winter, are very great, and it also happens that there are very few long-distance commercial routes which pass near to the pole, but it is useful to have the ideal theoretical conditions as a basis for analysing any available experimental data.

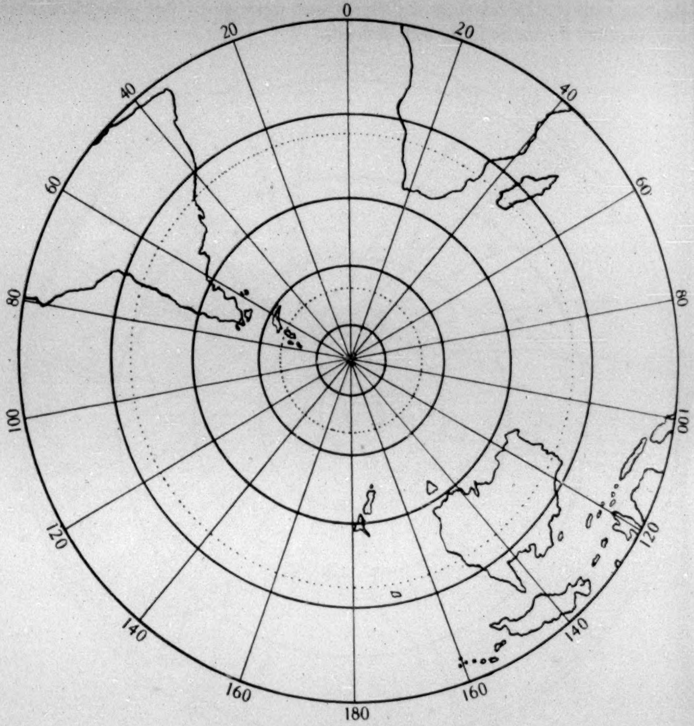


Figure 12. Southern hemisphere.

In conclusion it will be seen that the polar charts give a much better idea of the density conditions over the polar caps than can be obtained from the Mercator projection, and although they are not so convenient they can be used by the method indicated above to study routes which cross the equator. They also give a better idea of the overall daylight attenuation for a route lying in more than one grade and should have a special application in the study of magnetic-storm effects.

REFERENCE

(1) G. Millington. *Proc. phys. Soc.* **44**, 580 (1932).

621.396.615.1

SOME EXPERIMENTS ON ELECTRONIC OSCILLATIONS

By W. A. LEYSHON, Ph.D., F.Inst.P.

Received October 23, 1934. *Read with demonstration December* 21, 1934.

ABSTRACT. The paper gives results of measurements made on the wave-length of the oscillations generated by small triodes using the positive-grid (Barkhausen-Kurz[(1)]) method, in a fixed oscillatory circuit which, in most of the cases studied, was formed by the valve electrodes and leads. It was found possible, with the grid current emission limited, to express the relation between the generated wave-length λ, the grid voltage v_g, and the plate voltage v_p, by a straight-line graph connecting $(v_g - \mu v_p)$ with λ, where $\mu \left[\left(\frac{dv_g}{dv_p}\right)_{\lambda \text{ const.}}\right]$ was a constant for a particular valve. The results show that μ is constant whatever the value of λ generated by the triode and associated circuit. It is suggested that triodes connected in the manner described might be used as oscillation wave-meters working over a limited range.

---

§ 1. INTRODUCTION

In some preliminary experiments on electronic oscillations generated by small receiving triodes of the bright-emitter type the effect of joining the electrode leads through a condenser of capacity 0·0005 μF. was noted. The effect was greatest when the condenser was connected to the plate and filament leads, close to the valve. In a particular experiment it was found that the galvanometer deflection when resonance occurred in the wave-meter was increased ten times by connecting the condenser in this way. A large, but not proportionate, increase in the plate current of the valve was noted at the same time. With the condenser in position, oscillations could be obtained with lower filament current than was sufficient without it, other circuit conditions remaining unchanged. A second condenser between the plate and the other end of the filament was found to be a further improvement.

The increased strength of oscillation due to these condensers was illustrated in an experiment made with a Mullard *R* valve. A small flash-lamp bulb was inserted as a resistance in the condenser circuit, figure 1. The conditions were as follows: $V_g = 104$ V., $i_g = 39$ mA., $v_p = 2$ V. With the bulb in, $D = 20$ divisions and $i_p = 120$ divisions. With the bulb removed, $D = 700$ divisions and $i_p = 280$ divisions. These results show that considerable current was flowing in the condenser circuit. With this method of connexion, observations were made of the variation of the generated wave-length with alteration of grid voltage, plate voltage, and filament emission.

For convenience, the various quantities referred to in the paper are represented by symbols as shown below:

v_g v_g is grid voltage and $\}$ measured from the negative end of the filament;
v_p v_p plate voltage
v_f v_f is the filament voltage;
i_g i_g the grid current;
Δi_g Δi_g the change in mean grid current due to oscillations;
i_p i_p the plate current;
i_f i_f the filament current;
D D the deflection of the galvanometer at resonance in the wave-meter circuit, and
λ λ the wave-length of the oscillations, while
μ $\mu = \left(\frac{dv_g}{dv_p}\right)_{\lambda\text{ const.}}$

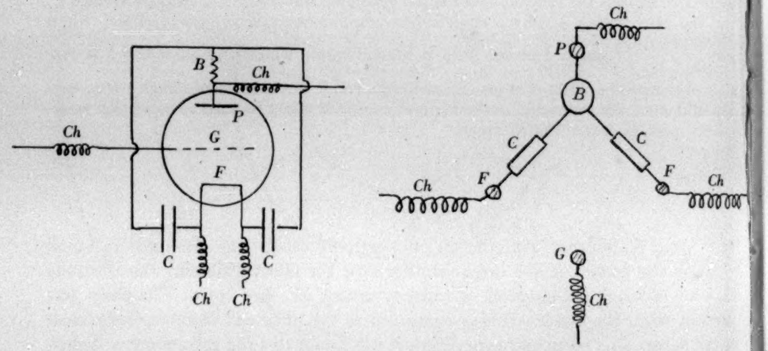


Figure 1. *P*, plate; *G*, grid; *F*, filament; *C*, *C*, condensers; *Ch*, *Ch*, chokes; *B*, flash-lamp bulb.

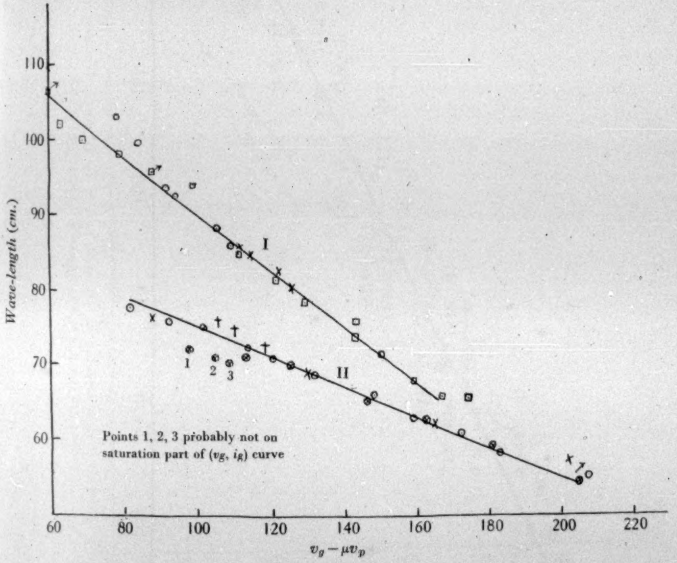
§ 2. DESCRIPTION OF APPARATUS AND EXPERIMENTAL RESULTS

The triodes used were two B.T.H. *R* valves, one, R_1, having had the pins removed; two Mullard *R* triodes, of different sizes; a *Q* valve, an L.S.I. valve, a valve *X* of *R* type of unknown manufacture, and a T.M.C. valve with a fixed Lecher-wire circuit attached to grid and anode, and bridged at the end remote from the valve by a condenser of large capacity. Thus experiments were made with triodes differing considerably from each other both in construction and in electrode-spacing.

The wave-meter used was like that described by Chapman(2). The coupling between the valve circuit and the wave-meter was loose enough for the plate current to be independent of the wave-meter bridge position.

It was found that the relation between grid voltage and wave-length for a particular value of plate voltage was a linear one, and that when the lines were drawn for different values of plate voltage they were parallel. Further, the relation between grid voltage, plate voltage and wave-length could be shown by a single straight line connecting $(v_g - \mu v_p)$ with λ, where μ had a different value for each valve tested.

Some curves showing the straight-line relationship between $(v_g - \mu v_p)$ and λ are shown in figures 2 and 3. The marked values of v_g and v_p indicated on the graph show the wide range of these quantities over which the relationship held in the case of the small Mullard valve, figure 3.



Points 1, 2, 3 probably not on saturation part of (v_g, i_g) curve

Figure 2. I, B.T.H. *R* valve (R_1) with pins removed, this valve contained a little gas, $\mu = 3·33$; ⊡ $v_p = -3$, $i_g = 17$–20 mA.; × $v_g = 100$, $v_p = -2 \rightarrow -8$, $i_g = 20$ mA.; ⊙ $v_g = 80$, $v_p = 0 \rightarrow -9$, $i_g = 20$ mA.; □↗ $v_p = -3$, $i_g = 10$ mA. II, B.T.H. *R* valve (R_2) with pins left on, $\mu = 2·5$; ⊙ $v_p = -3$, $i_g = 20 \rightarrow 34$ mA., blocking condensers 0·0003 mF.; †, × $v_p = -6$, $i_g = 15 \rightarrow 34$ mA., blocking condensers 0·001 mF.; ⊗ $v_p = -11$, $i_g = 16 \rightarrow 30$ mA., blocking condensers 0·0003 mF.; ↗ $v_p = -20$, $i_g = 40$ mA., blocking condensers 0·0003 mF.

A further experiment was made with the valve *X* by means of the circuit of figure 1 without the flash-lamp bulb resistance and with rather long choke coils, on which it was found that there were standing waves of quite considerable amplitude. The circuit was then set up again, the choke coils being omitted completely. The grid lead to the milliammeter was made as short and straight as possible,

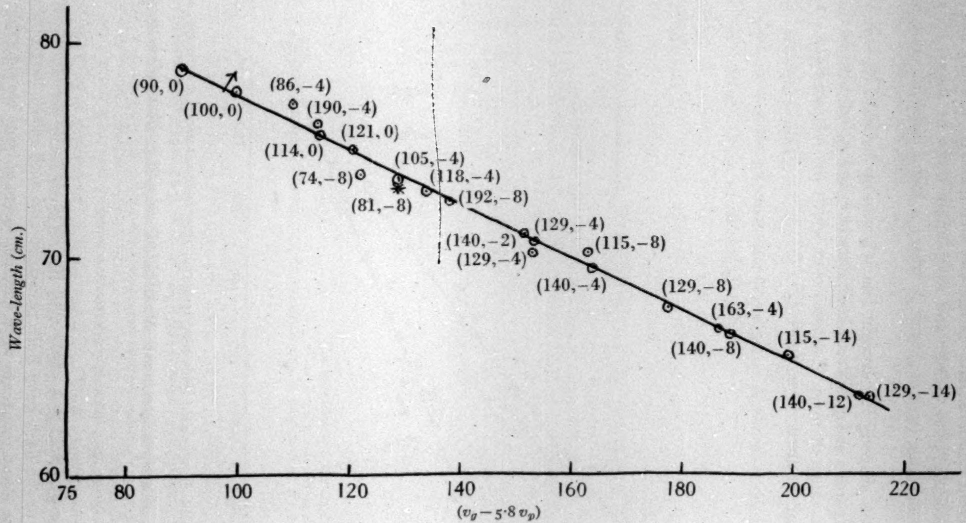


Figure 3. Mullard *R* valve working on saturation part of (v_g, i_g) curve. Blocking condensers 0·0003 mF. except at * (0·0001 mF.) and ↗ (0·001 mF.).

as also was the anode lead. The latter was run parallel and close to the filament leads. The valve oscillated over about the same range of voltages, but the slope of the $(v_g, \lambda)_{v_p \text{ const.}}$ curve was less. In figure 4 are shown curves connecting $(v_g - \mu v_p)$ with λ for these conditions. The value of $\mu \left[\left(\frac{dv_g}{dv_p}\right)_{\lambda \text{ const.}}\right]$ was, as would be

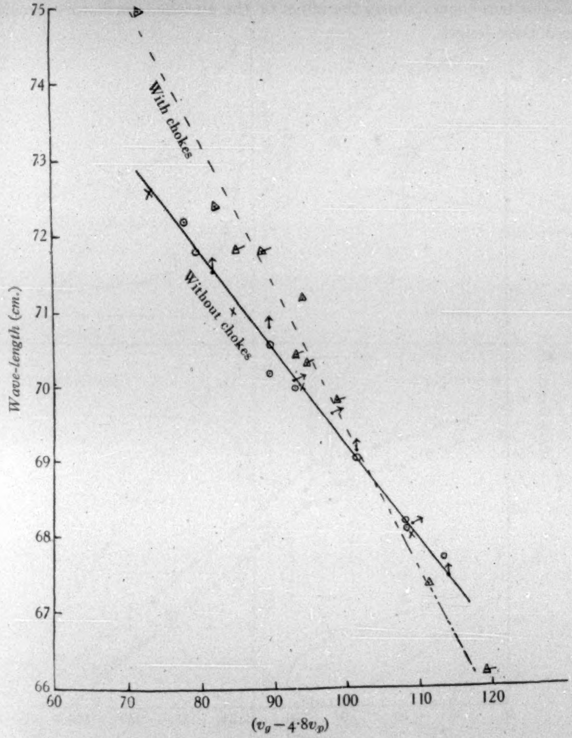


Figure 4. Valve *X*.
× $v_p = 0$; ⊙ $v_p = -2$; ↑ $v_p = -4\cdot5$; ↗ $v_p = -6\cdot5$, no chokes.
◬ $v_p = 0$; ⊿ $v_p = -6$, with long chokes.

expected, unaltered by changing the external circuit. This experiment shows how the change in wave-length resulting from a given alteration of electrode potential is diminished by reducing the effective resistance of the attached oscillatory circuit.

The straight-line relationship between v_g and λ was still found to hold, and the slope of the line was found to alter at a particular value of v_g dependent on the filament emission for a T.M.C. valve, with fixed Lecher-wire circuit attached to the plate and grid terminals (see figure 5). It was further noted that the condenser at the end of the fixed Lecher-wire system was definitely a potential node for smaller values of v_g, but as the value of v_g was raised it ceased to be so. In figure 5 is also shown a (v_g, λ) curve obtained with the circuit of figure 1, except that one condenser only connected plate and filament and *B* was omitted. The difference between the two curves shows the effect of the attached oscillatory circuit on the generated wave-length.

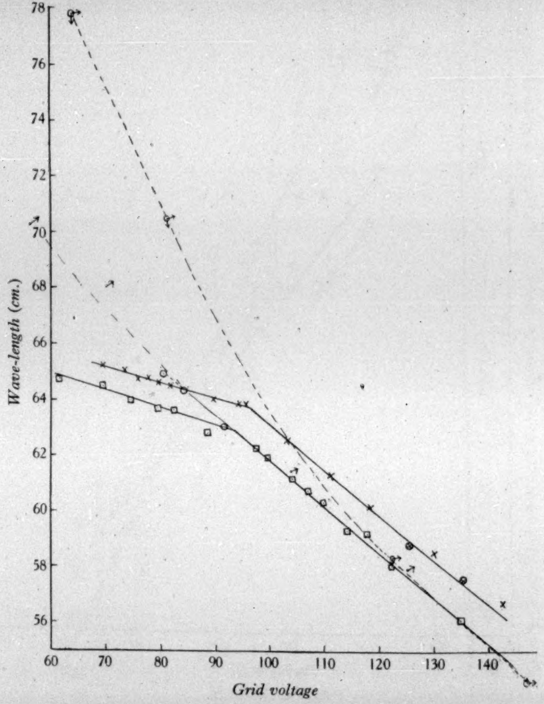


Figure 5. T.M.C. valve with fixed Lecher-wire circuit between grid and plate; ⊡ $v_p = -2{\cdot}05$, $i_g = 27$ mA.; ⊙ $v_p = -2{\cdot}6$, $i_g = 12{\cdot}75$ mA.; × $v_p = 0$, $i_g = 27$ mA.; ⊗ from graphs of v_g and position of λ meter bridge for max. deflection; ⊕→ calculated by Scheibe's method ($v_p = -2{\cdot}05$); ↗ $v_p = -2$, condenser between plate and $F-$ (August 1933 and August 1934).

In obtaining results with the circuit of figure 1 but with *B* omitted, i_p and *D* (for resonance in the wave-meter circuit) were noted, the position of the valve with reference to the wave-meter wires being left unaltered. Curves connecting v_g with i_p and v_g with *D* for a particular value of v_p are given in figure 6 for four different triodes.

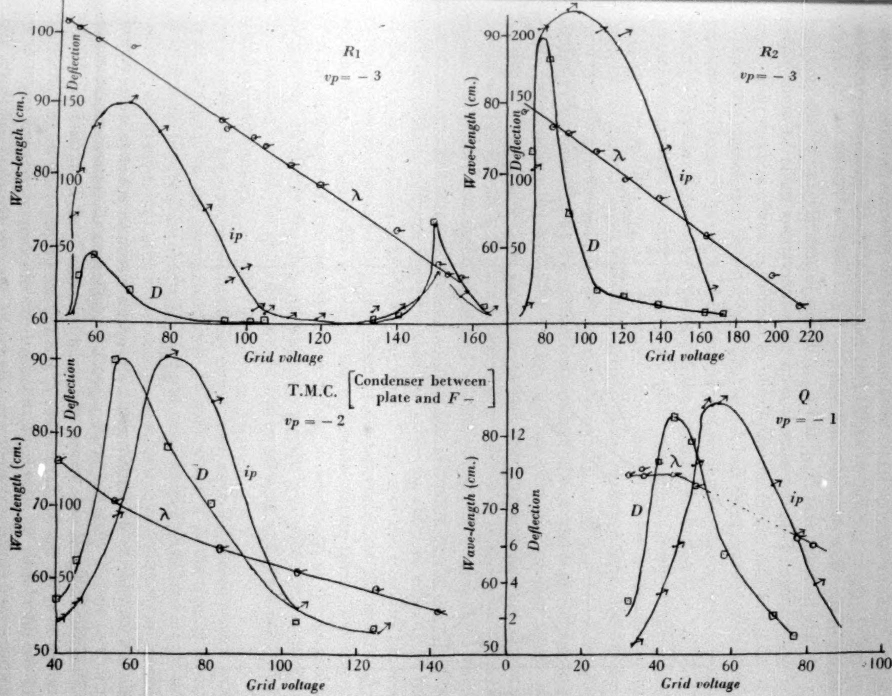


Figure 6 (July 1933).

The form of these curves suggests very strongly that some kind of tuning is taking place as the grid voltage is altered. They may be compared with the curve recently given by Gill and Donaldson(4) for a triode used as a receiver of ultra-short waves. In the present case it is to be supposed that the potential differences impressed on the electrodes are due to oscillations maintained in an oscillatory circuit consisting of the valve electrodes and leads.

It will be noted that $D_{\text{max.}}$ and $i_{p\,\text{max.}}$ do not in general occur for the same value of grid voltage. This corresponds with the observations made by other workers to the effect that maximum amplitude of oscillation and maximum plate current do not coincide as the length of a Lecher-wire circuit attached to the generating triode is altered while the static potentials applied to the electrodes remain unchanged. In figure 7 is shown the variation with v_g of D, i_p and Δi_g for two values of v_p for the Mullard valve; i_p and Δi_g are proportional to each other and of the same order of magnitude. Evidently in this case a certain fraction of the constant total emission current was alternately taken by the plate and by the grid. The variations in i_g were measured by means of a balanced galvanometer in the grid circuit.

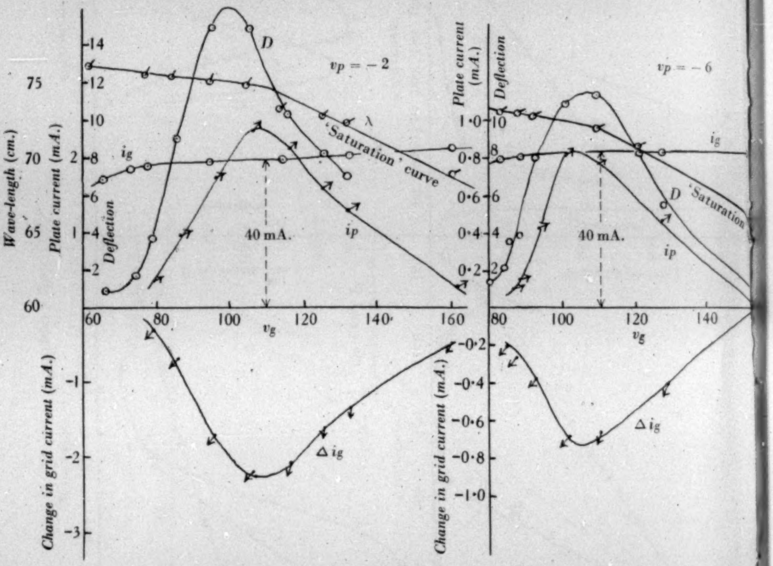


Figure 7.

In almost all cases in which i_g was recorded simultaneously with λ, a change in slope of the $(v_g, \lambda)_{vp\,\text{const.}}$ curve was observed in the neighbourhood of the top bend of the (v_g, i_g) curve, and therefore usually in the neighbourhood of the maximum amplitude of oscillation. The abrupt changes in slope of the (v_g, λ) curves shown in figure 5 for the T.M.C. valve coincided with the points of maximum oscillation, which in this case occurred for values of v_g considerably greater than that required to draw the saturation current from the filament. This triode contained a certain amount of gas.

§ 3. NOTES ON THE RESULTS

After the relationship between $(v_g - \mu v_p)$ and λ had been established, a paper appeared by Clavier(3) in which it was stated that for the frequency most strongly maintained by the micro-ray valves used in the Lympne-St. Inglevert 17 cm.-wave service, a linear relation between grid and plate voltages held over a wide range of values. The results given in the present paper show that this linear relationship holds for the very different oscillation arrangements described here, and furthermore that it holds not only for the frequency most strongly maintained but for *any* frequency generated by the system comprising the tube and associated circuit.

A linear relation between grid voltage and wave-length was also found by Grechowa(6) for a double-tube arrangement with fixed oscillatory circuit attached, but the effect of alteration of plate voltage was not shown as in the present experiments. The effect of an alteration of grid or plate voltage on the generated wavelength may be considered as twofold. Firstly, by alteration of the quantity and static distribution of space charge within the tube, the effective capacity of the electrodes may be altered. Secondly, the time of transit of the electrons between the electrodes is altered so that the phase of the current at the electrode under consideration is altered with respect to the oscillating voltage there. The change of phase is, of course, accompanied by a change in amplitude of the generated oscillations. If neither the filament-grid space nor the grid-plate space is carrying its saturation current the first effect would not be marked, and the second effect would be predominant.

In order to see how experimentally determined wave-lengths compared with those calculated on the assumption that $\lambda = 2cT$, where T is the time of transit of an electron between the electrodes in the absence of oscillation, and $c = 3 \times 10^{10}$ cm./sec., the dimensions of the valves used were measured as accurately as possible (i) by using a cathetometer and (ii) by projecting enlarged images on a screen by means of a projecting lantern. The electrodes of commercial valves of the type used are not likely to be of exact cylindrical symmetry, but the particular valves chosen approximated to this condition.

In the case of the triode R_2 the experimental results for the shortest values of λ fitted almost exactly the (v_g, λ) curves obtained by calculation of the electron transit time between filament and turning-point on the assumption that neither space charge nor oscillations influenced this time.

In the case of the other valves such exact coincidence was not found. However, in all cases, except that of the Mullard valve, the experimental values of λ were shorter than those calculated from $\lambda = 2cT$ according to the method given by Scheibe⁽⁷⁾.

Various authors have noted that the effect of increasing space charge as well as of increasing amplitude of oscillation is to shorten the wave-length of the electronic oscillations generated by a triode.

It may be concluded that the straight-line relationship between $(v_g - \mu v_p)$ and λ, for oscillations generated in a suitable fixed circuit by a positive-grid triode, has been established experimentally. A quantitative explanation of the relationship has not been found. However, lack of agreement between the measured wave-lengths and those found from calculated transit times must be ascribed to the disturbing factors not allowed for in the wave-length calculations, such as (i) the quantity of space charge and its distribution; (ii) the amplitude of oscillation of the electrode potentials; and (iii) the phase differences between these potentials and the currents to the electrodes.

The definiteness and repeatability of the results of the wave-length determinations with fixed circuit conditions and the linear dependence of λ on the applied potentials suggest that a triode connected in the simple way described might be used as an oscillation wave-meter over a limited range; moreover, the amplitude of oscillation could be altered by adjustment of the filament current without appreciable change in the wave-length, provided that v_g was sufficiently great.

§ 4. ACKNOWLEDGMENTS

The experimental work described in this paper was carried out mainly in the Physics Department of the London (R.F.H.) School of Medicine for Women; it was completed in the Electrical Laboratory at Oxford, with the kind permission of Prof. J. S. Townsend.

I should like to express my thanks to Dr W. H. Eccles for the interest he has taken in the work, and to Mr E. W. B. Gill for helpful suggestions made while I was working in the Electrical Laboratory.

REFERENCES

(1) BARKHAUSEN H. and KURZ, K. *Phys. Z.* **21**, 1 (1920); E. W. B. GILL and J. H. MORRELL. *Phil. Mag.* **44**, 161 (1922).
(2) CHAPMAN, F. W. *Exp. Wireless*, **9**, No. 108, 500 (1932).
(3) CLAVIER, A. *L'Onde élect.* **13**, No. 147, 101 (1934).
(4) GILL, E. W. B. and DONALDSON, R. H. *Phil. Mag.* **15**, 1177 (1933).
(6) GRECHOWA, M. T. *Z. Phys.* **35**, 50 and 59 (1925); **38**, 621 (1926).
(7) SCHEIBE, A. *Ann. Phys.*, Lpz., **73**, 54 (1924).

541.183.26

MODELS OF THE SUPERPOSITION AND INTERPENETRATION OF COMPONENTS IN GAS MIXTURES ADSORBED UPON THERMIONIC, PHOTOELECTRIC, AND CATALYTIC SURFACES:

PART I, PRINCIPLES

By M. C. JOHNSON, M.A., D.Sc., Physics Department,
Birmingham University

Received October 16, 1934, *and in revised form November* 14, 1934.
Read in title December 21, 1934.

ABSTRACT. Models of intermolecular attraction and repulsion are developed to represent the penetration of one adsorbed substance through another. Such interchanges between two or more components are important in view of the complex surface structures required for control of thermionic or photoelectric emission, and for the many physical and chemical experiments in which a metal is covered with a thin deposit of another metal over an intermediate gas layer; the latter may reach its final equilibrium between the metals after migration from above or below. Other relationships between successive adsorbed materials range from the protection of a layer by inert gas, to the disruption of a catalyst by penetration of reacting vapour. The concepts of bulk diffusion and electrolysis, applied successfully to comparatively thick oxides, etc., require supplementing when the exchanges take place between two layers each of single atomic thickness. Accordingly we adopt simplified models to enquire what properties a molecule must possess if it is either to penetrate and replace, or to remain superposed upon and stabilize, a previously adsorbed group of molecules. The range of models is chosen to cover certain main features of oxygen, hydrogen, caesium, and inert gases, adsorbed in dissociated, ionic, and other states upon a tungsten surface. The enquiry yields some of the conditions whereby any given metal-gas sequence is able to rearrange its layers into a stable equilibrium independent of the order in which the components were originally deposited. In successive approximations we consider first the adhesion of given molecular models to a clean surface, and to the same surface when screened by a previous deposit in the form of a complete sheet. Distribution of the previous deposit is next taken into account as giving access to the solid through interstices between inert or repelling atoms and ions. The restriction to immobility of the first deposit is then removed; the energy required for lateral migration is also considered for a second layer upon a first layer, and for the first layer loaded by the second layer. The work done in penetration of one layer by another then becomes determinate for the given models. Finally the modifications needed for fissures and edges on the solid lattice are briefly discussed, for application to the catalytic activity of isolated areas and grains.

---

§ 1. INTRODUCTION

THE production of thermionic filaments capable of large and lasting emission at low temperatures(1), and of photocells capable of a required sensitivity and spectral range(2), depends essentially on the formation of surface layers which are often very complex; they may include one or more gases and metallic vapours deposited, modified by discharge, and required to remain stable during long life exposed to the residuals of a high vacuum or to the impurities of an inert atmosphere. The attainment of equilibrium ensuring such stability may involve one or more exchanges of position by interpenetration between one layer and another.

The mechanics of adsorption(3) presented by such assemblages is not susceptible of exact treatment. We suggest, however, that by constructing very simple models, i.e. by assuming certain laws of force and dimensions of spheres of molecular influence, it is possible to determine within specified limits the extent to which two components bearing at least some of the properties of known gases may be expected to interpenetrate; this distinguishes ultimately between mixtures which must be deposited on a surface in the order finally required, and mixtures which may be left to arrange themselves.

The present paper is concerned with outlining the method for all likely substances rather than pursuing detail of any one of them; the terms "caesium," "oxygen," etc., are only used to describe highly conventionalized constructions chosen as caesium-like, oxygen-like, etc., from one or two representative tendencies.

§ 2. ATTRACTION OF A MOLECULE OUTSIDE AN ALREADY ADSORBED LAYER

Method. The chance of a molecule from the gaseous phase penetrating or displacing any outer layer on which it impinges depends on (*a*) the time during which it can remain attached before re-evaporation, (*b*) the interstices of the outer layer, and (*c*) the lateral movements of which the atoms of the outer layers are capable during that attachment. Since (*a*) is known to vary exponentially with temperature and with the potential energy of the attachment, its range of variation is obtainable by determining the difference between energy of adhesion to a clean surface and energy of adhesion to the same surface if the latter is already covered by a screening layer of adsorbed gas or vapour. Before discussing in § 3 the structure of such a screen, we consider the simpler case in which it is regarded as a locally plane and complete sheet, whose thickness d' is equal to the supposed diameter of the adsorbed particle.

In the simplest case the attraction of impinging gas to the solid is unaltered in character by the interposed screen, but merely diminished in magnitude owing to the gas being prevented from closest approach to the solid. In more complex cases the screen itself adds its own attraction to that transmitted through it from the solid, as in the cementing of an alkali to tungsten by a gas layer. In extreme cases the screen or even the surface layers of the solid become detached by attraction to incoming gas, as in the formation of WO_3, the action of CO on Ni, etc. A further complexity is introduced if the screen modifies qualitatively and not only quantitatively the forces between solid and gas. For Van der Waals's forces this modification is probably slight. It may become important, however, if the screen is held by a valency bond to the surface, as in the so-called chemisorption of O_1 and H_1; if this bond is saturated by the first layer, similar particles approaching from the gas may no longer feel even the weakened residual attraction appropriate to their greater distance from the solid. For instance if Ni has caused the formation of a layer of H_1 by surface dissociation, subsequently impinging H_2 forms a second but probably undissociated layer by Van der Waals's attraction only, and any H_1 generated in the gas, e.g. by discharges, will also have to adsorb by Van der Waals's attraction. A second layer of mobile oxygen upon a first layer of rigid oxygen covering a metal filament is probably also to be explained in this way.

Starting with the simplest case of the plane, inert, non-modifying layer, if the laws of variation of attraction and repulsion with distance can be reliably assumed, the action of this interposed screen can be represented by simply displacing outwards the repulsions and retaining the same attractions as were felt at corresponding distances from a clean surface. We combine the attraction appropriate to a distance d from the solid with the repulsion appropriate to a distance $d-d'$, when d' is the thickness of the screen. This type of model is only strictly valid if the atoms of the first adsorbed layer are similar in hardness to the solid itself. d, d'

Let the potential energy E_1 of an impinging particle at a distance d from the bare solid be given by E_1

$$E_1 = \lambda d^{-n} - \mu d^{-m} \qquad \text{......(1)},$$

where n and m represent laws of decrease of repulsive and attractive energy with respect to the solid, and λ and μ are constants which must be chosen, for any n and m, to enable the position of equilibrium under all the forces to occur at a correct distance from the surface. Then for the potential energy E_2 of attraction of the same atom to an already gas-covered surface, this approximation gives n, m λ, μ E_2

$$E_2 = \lambda (d-d')^{-n} - \mu d^{-m} \qquad \text{......(2)}.$$

Graphs of E_1 and E_2 plotted against d illustrate the effect of the screening layer in causing the potential minimum, representing position of equilibrium in adsorption, to become rapidly shallower as it is displaced outwards, figures 1, 2 and 3.

Values of the constants. The effects of d', and the distance at which the first potential minimum occurs, and also the distribution of adsorbed atoms to be considered later, require some convention to be adopted as to atomic dimensions in gases, vapours, and solids.

The size of any atom varies greatly according as it is defined through (*a*) gas-collision target, (*b*) X-ray crystallographic packing, (*c*) maximum density in charge distribution of an outer electron shell. Since O_1 and H_1 are inaccessible by (*a*), and (*b*) yields mainly ionic states, a complete set of radii is only obtainable through (*c*) and will appear excessive or deficient, according to the nature of the surrounding

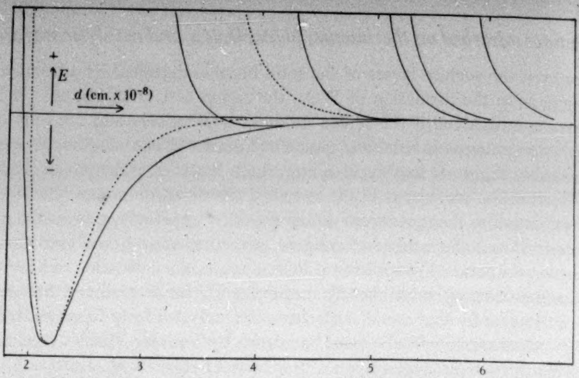


Figure 1. Model 1, $n=12$, $m=6$ ———. Model 3, $n=12$, $m=11$ ········.

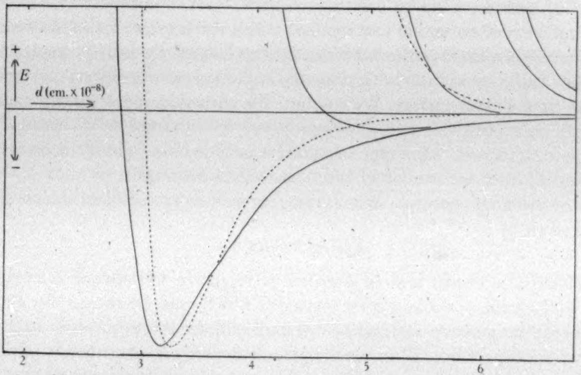


Figure 2. Model 2, $n=12$, $m=6$ ———. Model 4, $n=12$, $m=11$ ········.

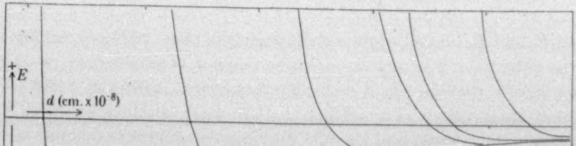


Figure 3. Model 5, $n=12$, $m=3$.

fields, when comparison with (*a*) or (*b*) arises. The definition of radii through atomic wave functions has been developed by Slater, Hartree, Pauling, and others; although Slater's comprehensive table$^{(4)}$ has been superseded for certain atoms, it fulfils best our need of exhibiting all atoms on a single comparable basis: accordingly we adopt the mean of his O_1 and H_1 as smallest possible model, and his Cs as largest possible model. These exaggerate both upper and lower limits, and his tungsten constitutes a too overlapping lattice. But for our purpose of emphasizing consequences of differences in size these exaggerations are an advantage rather than otherwise in this general survey. Our diagrams differ radically from those employed in crystal analysis, where overlapping and separation are both avoided by confining definitions of radii to (*b*) alone, and where the structure is in equilibrium, under forces not necessarily those of a monomolecular adsorbed layer, so that the problem of penetrability does not arise.

Table 1

	H_1	O_1	He	H_2	A	O_2	Cs	W	Cs^+	$O^{-(-)}$	
$\frac{1}{2}d'=$	0·5	0·5	1·0	1·4	1·5	1·8	4·25	2·7	1·75	1·35	$\times 10^{-8}$ cm.

In any model of an interface the distance between centres of adsorbed and adsorbing particles might be approximately the sum of these structural radii, or if the adsorbate packs as if taking part in certain types of chemical reaction the distances might be compressed to the order of $\frac{2}{3}$ the sum of the radii. To compare such alternatives we choose values of λ and μ which will give minima in the potential curve at $(d'/2+\bar{d}'/2)$, models 2 and 4, and also at $\frac{2}{3}\,(d'/2+\bar{d}'/2)$, models 1 and 3, where $\bar{d}'$ is the diameter of the atom in the solid. The repulsive index *n* we take as always 12, and it is not likely to be smaller except possibly for the soft alkalis. *m* is known to be about 6 for Van der Waals's forces, models 2 and 1; 3 for dipole forces, model 5; and much higher for the homopolar bond. Since this latter can be imitated by an induced polarization involving the square of a dipole moment, and the dipole moment may itself vary with distance up to a fourth power, the greatest steepness in attractive potential can be represented suitably by giving to *m* the value 11, models 4 and 3. Actually these forces may decrease exponentially with distance, but to a considerable approximation and with greater computational facility a power model suffices.

$\bar{d}'$

Figures 1, 2 and 3 show the equilibrium of a second adsorbed layer in comparison with that of a first, calculated from the above values inserted in equations (1) and (2). Each group of shallow curves represents the combined attractive and repulsive energy of an impinging atom when the surface is covered with the several inert-model screening layers; the single deep curve of each group represents the corresponding values of the quantities if the surface were bare. Thus for these simple models it it possible to see at once how much less work has to be done by the solid in evaporating a second layer than in evaporating the first layer, as measured by the decreasing depth of the minima; we also see the dependence of this decrease upon the type of attraction exercised, as represented by its particular laws of force. While sufficient range of variation with distance is covered by the several models, the scale of ordinates will differ in different cases served possibly by similar models; e.g. the depth of the trough in the curve for O_1 and H_1 is of the order of 10^5 calories per gram-molecule, but for He at the other extreme it is of the order of 10^2.

E_1' E_2' *Results.* Let E_1' be the energy of adsorption of the first layer on the bare surface, i.e. the work to be done in evaporating the first gas. Let E_2' be the corresponding value for the second layer on the gas-screened surface. The values of E_2'/E_1' are tabulated in table 2 from the several graphs, showing the reduction of adsorptive power by the inert screens.

Table 2

d' (cm. × 10^{-8})	1·0	2·0	2·8	3·0	3·6	Adsorption model
E_2'/E_1'	0·29	0·13	0·08	0·06	0·04	5
	0·07	0·01	0·005			2
	0·03	0·005				1
	<0·005					4
	<0·005					3

We proceed to indicate briefly the application of this table.

(*a*) One of the largest known values of E_1' is that for O_1 on W, estimated by Langmuir(5) to be about 160,000 calories per gram-atom. The models 4 and 3 are the most appropriate for such an adsorption which has the strength of a chemical attraction of extremely short range; from the above table it is seen that only 800 cal. could remain as the maximum possible energy of adsorption outside a close screen whose thickness is that of the smallest possible atom. Independently of whether such screen does itself saturate the available valency of the metal, we have from this figure a reason why no second layer of O_1 is found by surface dissociation, and also a reason for expecting O_1 or H_1 present in a gaseous discharge to be only weakly adsorbed, if the W is already covered by a packed layer similar to this model.

(*b*) The next largest values of E_1' belong to metallic vapours, e.g. 65,000 cal., is estimated in Langmuir's 6) experiments with Cs on W. The forces involved are not necessarily of the shortest range, and models 2 and 1 are allowable, giving the maximum energy of binding of a second layer as 4500 cal. through a model screen corresponding to O_1 or H_1, 650 through He, and 300 through diatomic gases.

(*c*) Van der Waals's forces provide heats of adsorption up to the order of 600 cal. only, and these will be reduced on the same scale to 400, and to less than 100 cal., through H_1, He, and diatomic layers.

(*d*) The longer range forces of model 5 are less drastically diminished by the gas screens, but in their case the value of E_1' is less initially. A strong dipolar molecule provides energy of surface adhesion of the order of 3000 cal., which will be reduced by the H_1, He, and diatomic layers to 900, 400, and 300 cal. per gm.-mol. respectively.

(*e*) With the forces of longest range, E_1' may be large if due to the attraction of an opposite ion in a polar crystal; at a metal surface an ion or a molecule of dipolar structure will experience the similar attraction to its image at a corresponding distance behind the surface. A screen of monomolecular thickness may double the distance of a layer while the energy of image attraction of gaseous ions falls only to half value; at the smallest atomic distances the mirror plane is an inadequate model, but the decrease of potential with distance will be closely linear outside the regions of chemical attraction.

Two standards of comparison must be used in applying these and similarly derived estimates of the diminished energy with which a second substance, or a second layer of the same substance, adheres outside a first adsorbed layer. Firstly take the attraction of the screen itself, which should be added to that which it transmits from the solid. For screens of inert gas, or of active gas whose valency is saturated, this attraction will be very small; in other important cases the attraction to the screen is of the same order as that of the screen itself to the solid, as when a diatomic molecule is formed from H_1 striking adsorbed H_1, or when CO_2 is formed at a Pt surface, or when multimolecular layers are formed from most vapours. The extreme case when atoms of the solid become detached has been mentioned. Qualitatively such processes may be represented as transitions from the troughs of figures 1, 2 and 3, to other troughs of potential energy with respect to isolated centres in the adjacent gaseous phase, as depicted by Gurney and by Fowler in the study of electrolysis. Since the energy in the final trough is often that of some well-known gaseous compound, comparison may be effected between this known energy and the energy of adsorption, the latter being less than the former if the reaction proceeds irreversibly.

The second standard of comparison is that of the vibrational energy levels in the adsorbed state; the energy of thermal equilibrium is of the order of 300 cal. per gm.-mol. at 100° K., 900 cal. at 300°, and 3000 cal. at 1000°. Hence comparison of these with the data derived above will decide, for any given model at a given temperature, the chance of a second layer adhering to a first layer long enough for completion of the gradual mutual penetrations which we discuss later.

In addition, the graphs indicate the distinction between rates of change of energy with distance for nearer and further layers; these determine the work done during any small normal displacement involved in lateral migration of adsorbed particles, and will be used below in estimating the penetration of layers by their mobilities.

§ 3. PENETRATION THROUGH INTERSTICES OF A RIGID LAYER

Distribution of an immobile inert adsorbate. We next remove the restriction by which the first adsorbed gas was regarded as a plane sheet completely withholding further gases from the solid. Let saturation of any layer be defined, for the present needs, as the state in which the maximum number of particles is present, consistent with their equilibrium, at a given temperature, in the fields due to neighbours and to the solid below. When a steady state has been reached, the surface density at saturation may either fall short of close packing or may exceed close packing, if the latter be defined as the number of particles which can be arranged with their diameters in contact. Figure 4 illustrates models in which this excess or deficiency affects a layer's penetrability towards later arrivals from the gas phase. In this diagram the outer electron shells of the metal are represented by the dotted circles, and the atoms adsorbed on face or edges are represented by the full circles.

In using the values of *d'* from table 1 both neutral and negatively charged oxygen atoms must be taken into account as well as the diatomic molecule. The accepted diameter in crystal compounds, $2{\cdot}7 \times 10^{-8}$ cm., refers to an ionic state, but although some writers have suggested this state for adsorbed oxygen, the proof by detection of evaporating ions would be exceedingly difficult; such evidence as exists indicates that the evaporated adsorbate is neutral, both in the case of O_1 and H_1. On the other hand the alkalis evaporate as positive ions, according to a well-known law, from a surface of large enough electronic work function, while some at least of their atoms behave as neutral when still adsorbed. It is not possible to distinguish sharply between stable existence as a charged ion in an adsorbed state and mere loss or gain of an electron in separating from the surface; development of the new theory of molecular orbitals will probably remodel such states as are incapable of permanent association with a fixed number of electrons, and distinction between ions and neutrals is merely an approach from two extremes between which any actual facts must lie.

Neutral atomic oxygen and hydrogen. If the number of gas atoms in a saturated layer on tungsten is limited chemically to a final proportion of one to each exposed atom of the metal, we must suppose that the components of a diatomic gas molecule, dissociating under surface forces, experience initial attraction to several neighbouring metal centres. Round each of the latter a spherical shell will represent the locus of a potential energy minimum such as was studied in § 2, and the position of an adsorbed atom in equilibrium on the crystal lattice will be the nearest point to the intersection of as many as possible of such shells. This position is marked with shaded circles on the face and top edge in figure 4, in contrast to the positions marked + which would satisfy similar valency considerations but would be unstable for small displacements unless the gas-tungsten complex possessed a moment resisting rotation. Since the heat of adsorption of O_1 on W, attributable to a very intimate bond, is so enormous, the models of figure 1 are appropriate, so that the radius of the sphere of equilibrium is even smaller than the bounding radius of the W atom, if the adsorbed particle is a normal neutral atom. Considerations not very different apply to H_1.

An important feature of any such small atomic model is that tungsten covered by the neutral gas screen will present, even at saturation, a surface with the greater part of the metal still exposed, though with its free valencies no longer fully available. Similar impinging gases can therefore penetrate in most places unimpeded to the W; but if the valency of the metal is already saturated, this will limit subsequent adsorption to a Van der Waals's mechanism which, though weak, does not restrict their final density of packing. Such additional O_2 and O_1 will not follow any lattice structure, and are shown on the lower edge of figure 4. Hence even when the screen considered in § 2 becomes locally of vanishing thickness, so that the results given in table 2 are only applicable at separated points, it is still necessary to regard the second oxygen particle in Langmuir's OOW as a diatomic molecule, penetrating easily between the large interstices of a strongly held layer of atoms but unable to dissociate. By similar mechanism we may explain the inhibition of hydrogen dissociation at an oxygen-covered surface.

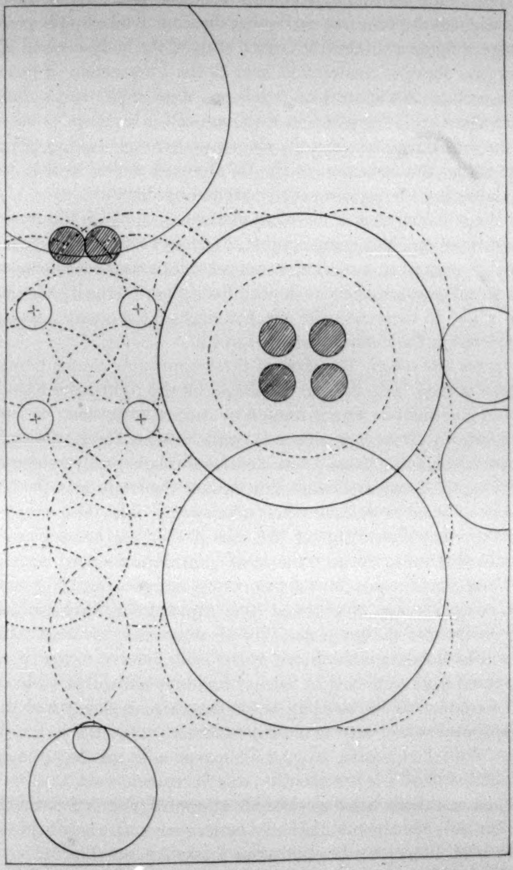


Figure 4.

The presence of O_2 or O_1 on W cannot be regarded as in stable equilibrium; the formation of volatile WO_3 when excess oxygen is admitted shows that further transition to a new grouping occurs, with a loosening of the bond from tungsten to tungsten.

Caesium. In contrast, the model representing extreme alkaline properties has a saturation density actually in excess of close packing if the neutral state is assumed; for any model approaching this extreme, penetrability to outside gases becomes unlikely except by the lateral displacements discussed later. If the adsorbed atoms crowd as closely into the W as into each other the Cs will take up the position shown at the top edge of figure 4; in this the central atom of the body-centred W lattice has the four oxygens above it clustered as near to the intersection of their minimum equipotential surfaces as their mutual repulsions allow, with one Cs atom above the group. Here again the oxygen-screen thickness of § 2 becomes vanishingly small, explaining the well-known fact that the Cs is more strongly held to OW than to W; since in our model the attraction of the O_1 atoms is added to that of the metal without the latter itself being seriously weakened by distance.

Van der Waals's adsorption of inert and diatomic particles is illustrated along the lower edge of figure 4. Not being subject to valency restriction, these and many metallic vapour layers when saturated do not present permanent interstices for penetration by external gases, and have to depend for such upon the displacements which we consider in § 4. In these cases the effective thickness of screen does not fall short of that considered in the models discussed in § 2.

Ions of oxygen and alkali. The sizes of these oppositely charged ions are not so unlike as those of their neutrals, and are shown on the right-hand edge in figure 4. The distribution of any one sign is limited by mutual repulsion.

The penetrability of any layer depends on the fields of force existing between its constituent particles, and it is only with neutral and spherically symmetrical atoms that this field can be considered negligible outside their boundary.

Distribution of immobile mutally repelling adsorbate: ions. The energy of mutual repulsion scatters any assemblage of like ions if they are not enclosed, and in a closed space would cause an appearance of enormous internal pressure on any boundary. Three consequences for the theory of penetration are to be noted: (i) The constituents of even a very unsaturated layer experiencing such mutual repulsion will tend to escape into the gas phase. The alkalis possess an inert core, therefore ions of Cs will be unrestrictedly driven apart, while ions of oxygen and hydrogen may be prevented from scattering by valency bonds to individual atoms of the solid. (ii) The electrostatic field surrounding an ion decreases so slowly with distance that it nowhere approaches zero in its lateral fluctuations over even a very sparse sheet of charged particles. Penetration by outside gases will involve passage through whatever potential barrier is presented by this fluctuating field, and cannot rely on interstices such as existed between neutrals of similar size and packing. (iii) Adsorbed ions can only remain in a stationary pattern when strongly held to individual atoms of the solid; otherwise they will show a gas-like mobility.

Distribution of dipoles. If the work function of a metal is not such as to alter the net charge of the adsorbed atoms, they may nevertheless suffer considerable distortion in the field of the surface, resulting in behaviour as dipoles towards neighbours. The mutual repulsive energy E between two like dipoles of moment μ reaches a maximum, under favourable orientation, given by

E, μ

$$E = \mu^2/d^3.$$

An estimate of μ is available since Langmuir considered the heat of adsorption of isolated Cs atoms on W was about 65,000 calories per gram-molecule, but for Cs in a saturated layer about 41,000 cal. The deficiency represents approximately 17×10^{-13} ergs per atom. If this is accounted for by mutual repulsions at the mean distance of 6×10^{-8} cm. used in figure 4, the repulsion of one particle by four similar neighbours implies that

$$\mu = 9{\cdot}6 \times 10^{-18} \text{ e.s.u.}$$

Conclusion of § 3. The dimensions and force laws of the models appropriate to neutral O_1 and H_1 allow gas molecules to penetrate freely through wide interstices in the first adsorbed layer; but as these models are also associated with saturation of the valency of the metal with respect to those gases, any second layer, though reaching the solid, is only weakly held there. The models corresponding to large alkali atoms and to diatomic molecules allow no such interstices in a saturated layer. Ions cannot be packed as densely as neutrals, but their scattering is greater for an inert than for an active core. Even a sparse ionic layer covers the whole surface with its potential barrier. The repulsion of distorted neutrals exerts a calculable but slighter effect on their spacing, and hence on the penetrability of a layer.

§4. PENETRATION BY DIFFERENTIAL MOBILITY OF LAYERS

Migration on clean surface. It is well known from experiment that the state of aggregation of adsorbed gases and vapours ranges from an almost complete freedom of migration over the surface, giving a two-dimensional gas, in some cases, to a rigid binding with either a characteristic spacing or the spacing of the underlying solid, giving a two-dimensional lattice, in others. Lennard-Jones(7) has initiated methods for calculating the energy which an adsorbed particle in a first layer must acquire in order to migrate from one cell to another on the surface of certain types of lattice. When a second layer is adsorbed upon the first layer in the manner we have discussed, the chance of the former penetrating the latter will obviously depend on its ability to migrate along the screen until it finds any of the interstices mentioned in § 3. Similarly, any migration of atoms in the screening layer itself will contribute towards opening such interstices.

Figures 5 and 6 show the components, in the plane of the surface and normal to it respectively, of such a change in position. In figure 5 the largest and smallest circles represent the boundaries of models used in § 3 for Cs and O_1; the dotted circles are sections of spheres whose radius is equal to the distance of the potential minimum in figure 1 from its attracting W centres. The length and direction of the horizontal arrow indicates the smallest movement by which an adsorbed atom, in equilibrium above the central W atom of the body-centred cube, can migrate to the top of a pass without actual evaporation; once the energy for this movement is obtained the atom can wander over an infinite network of such paths without needing further gain except to counteract losses by collision. An oblique movement, ending at the head of the vertical arrow, would require the surmounting of a higher pass and allow a wider choice of subsequent migration routes. The main lateral component of migration is derivable graphically or algebraically from the change in distance

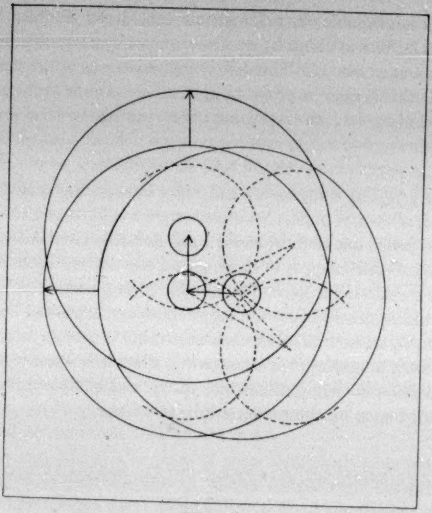


Figure 5.

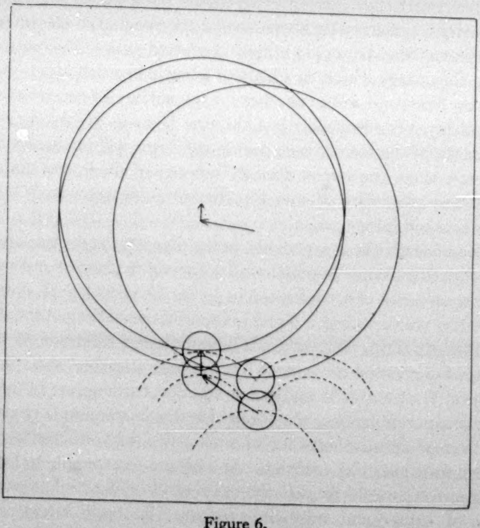


Figure 6.

relative to central and right-hand W atoms, indicated by the dotted straight lines.

In the plotting of the normal components, figure 6, the dotted circles around the W atoms are supplemented by smaller circles which trace the intersection of pairs of the equipotential spheres represented by the former; the innermost pair of O_1 positions show migration over these smaller circles, corresponding to the horizontal arrow of figure 5, and the outermost pair show the further migration at right angles, along the vertical arrow of figure 5. Three-dimensional components of displacement are taken from these diagrams, and the distance from each W atom at the top of the pass between lattice cells is compared with that at the trough in the centre of a cell. Referring then to the appropriate model in § 2, the work done in removal from trough to pass is calculated to the accuracy needed.

We find thus that if the energy of binding of the O_1 is 160,000 cal., the energy needed to mount the nearest pass, and hence to migrate over the surface, is nearly 90,000 cal., or more than 50 per cent. For Cs, 12,500 out of 65,000 cal. are needed. It is not unexpected that the smallest adsorbed particle, capable of penetrating almost through the solid lattice, should be so much less mobile than the large Cs; the estimate for the latter is close to that found in Langmuir's thermionic experiment with Cs on W, where the energy of migration is 21 per cent of the heat of adsorption, compared with the 19 per cent calculated from our model. Particles held by the much smaller Van der Waals's forces will have a much greater mobility.

Adsorbed ions are attracted to images in a plane surface; at such close distances the ideal plane is not an accurate localization, but an approximation from figure 4 places the work of migration of an ion between 5 per cent and 10 per cent of its work of evaporation.

Migration on gas-covered surface. Combining the results of the preceding section with those of § 2, it becomes evident that the great decrease of attraction which we found outside a screening layer must render all above that layer extremely mobile with respect to the underlying solid; not only is the residual attraction weak, but the graphs of § 2 show that the work done in migrating is further decreased by the flattening of the potential curves at great distances, until the path from trough to pass in a second layer becomes almost an equipotential. Only when the screening atoms themselves act as attracting centres will the second layer approach rigidity, and then only such measure of rigidity as was also possessed by the first layer.

Penetration due to migration of upper layer. The model of greatest rigidity, corresponding to neutral O_1, was also in § 3 that of sparsest distribution; hence, apart from inhibition of penetration by lateral repulsion as described below, any atom striking any part of the oxygenated surface needs only the slightest displacement to reach the underlying metal.

Penetration due to migration of lower layer. At the other extreme the very large and overlapping Cs atoms will not allow any penetration from outside unless they are pushed apart, which action we found to require the large energy of 12,500 calories per gram-molecule, greater than the total adsorption energy of any substance held by Van der Waals's forces alone, and greater than the diminished energy (§ 2) of even the strongest adsorption outside the Cs. It is experimentally known that oxygen adsorbed upon an alkaline-earth metal on tungsten succeeds slowly in penetrating and reversing the order, e.g. from WBaO to WOBa at about 1300° K. If these particles are neutral and have any likeness to the models we have used, our investigation suggests that the outer layer migrates easily over the lower one, but that until the lower is forced by the heating of the tungsten to migrate, mutual interpenetration cannot set in. A similar effect probably underlies the diffusion outwards of a layer of alkali from a bulk alkali already covered with a gas layer. Observation of the temperature variation of these exchanges would provide, on our theory, information as to the mobility of the lower rather than of the upper layer.

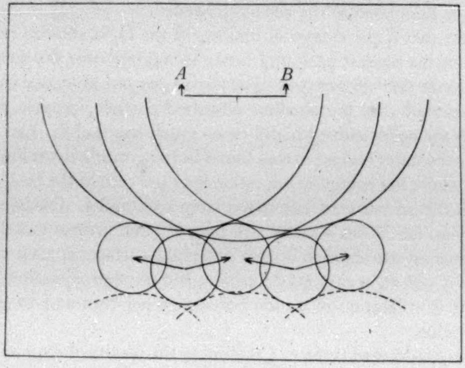


Figure 7.

Inhibition of mobility of lower layer by comparative rigidity of upper layer. This is illustrated in figure 7, with respect to the extreme example of a layer of helium models adsorbed on the metal at low temperature, with Cs deposited subsequently. An unprotected layer of He would only remain unevaporated at a very low temperature, and its lateral mobility would be very complete. But to achieve the migration shown by the arrows in figure 7 the protected or loaded He will have also to gain the energy needed to displace the Cs atom normally through the distance *A*, *B*. The large energy of migration of Cs found at the beginning of the present section is greatly reduced outside the He screen, but may still exceed the small migration energy of the He or even the energy required for the latter's evaporation. Such an immobilizing of an inner by an outer layer occurs in the trapping of molecules during deposit of sputtered and evaporated metal in the presence of gas, and is responsible for many of the electrical and optical properties of films so formed.

In contrast to the penetration due to migration of the lower layer, interpenetration of this double layer will depend mainly upon mobility being attained by the upper atoms, and temperature coefficients must be interpreted accordingly.

Inhibition of penetration by lateral repulsion. If the first layer or screen is no longer inert towards impinging gas, the initial requirement (§ 2) for penetration has the greater chance of fulfilment, through the lengthened interval before re-evaporation of the second layer; but an opposing tendency is also introduced if screen and gas molecule form together a complex which possesses preferential orientation with respect to (*a*) neighbours, or (*b*) the underlying solid. In some strong bindings of gas to screen, access of the former to the solid would be opposed as involving work of rotation of the combined structure. In case (*a*), this work is needed to overcome the lateral repulsion considered in § 3, if the layer is closely packed. In case (*b*), even without close packing, work has to be done in interpenetration of the two layers if attraction of the gas is exclusively to the screen, since distortion may then repel the outer atom from the solid. Such a system approaches the strictly oriented organic adsorption on liquids, studied by Adam and others, the nearer and further portions of a long chain molecule corresponding to the lower and upper atoms of our double layer. The directional properties of some types of valency bond may also appear in certain adsorptions, reinforcing the tendency for the second layer to remain outside the first and not penetrate to the solid.

Two-phase single layer of two components. When penetration has been achieved, there may have occurred a complete exchange of position between upper and lower layers, or, if one component is not much more strongly bound to the surface than the other, the two may form a mixed layer but still one of monomolecular thickness. In many cases, owing to discrepancy in size of atoms, etc., two such coplanar components will not be of the same mobility; thus arises the phenomenon of a two-phase adsorbed layer, conforming neither to the homogeneous two-dimensional gas nor to the two-dimensional lattice.

It is possible to illustrate (figure 8) the restriction which an immobile component may exercise on the migration of a companion. The closed contours round the summit of each lattice atom limit the regions over which a certain adsorbed particle can wander at a given temperature; the solid has also adsorbed a few isolated Cs and O_1 particles, whose rigidity was found to be considerable at moderate temperatures. The path of the migrating particle in the absence of any second component is restricted only to the shaded area between these equipotential contours; but where this area is covered by comparatively immobile Cs or O_1, freedom of movement in their directions is curtailed. A very sparse distribution may even block several adjacent channels and keep any one migrating atom to a small region of surface. Wherever any irregularity of distribution in the underlying solid allows a local opening of path, otherwise closed by the interspersed O_1 or Cs, a sudden increase in migration takes place; the picture thus adapts itself to the phenomena of lattice edges and fissures to be discussed in § 5.

Conclusion of § 4. We have shown, by analysing the components of displacement of given models, that the probability of an adsorbed atom being laterally mobile or rigidly attached will depend on its size and on the force laws of its adsorption bond. Selected examples of each of these dependencies range from rigid oxygen through semi-rigid alkalis to very mobile diatomic and inert gases and ions and the greatest mobility of all substances on gas-covered surfaces. Penetration of a second layer through a first layer is in certain cases determined by mobility of the outer component, in others by that of the inner, but may be inhibited by the loading of one atom by another and by the lateral repulsion between complex molecular structures. If a penetration is effected, a single layer of more than one component exhibits the state of aggregation of a fluid whose movement is restricted to definite channels.

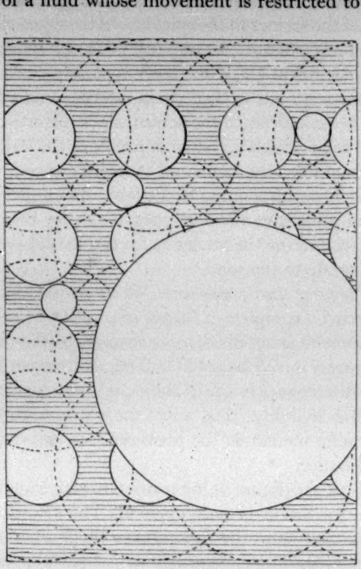


Figure 8.

§ 5. PENETRATION AT EDGES AND FISSURES IN A SOLID SURFACE

It is well known that irregularities in the microcrystalline structure of solids play a large part in the catalytic activity of their surfaces, and also in the extent to which impurities from both inside and outside a metal become enabled to modify its thermionic and photoelectric emission. If a group of surface atoms becomes partially isolated by the non-uniform distribution of its neighbours, its valencies must be less saturated than if it were completely surrounded, so that a denser clustering of any single adsorbate will be liable to occur in the neighbourhood; further, if two components are present in the gas phase, the increased probability of their adsorption in juxtaposition to each other at such points, edges, and corners, will increase locally the catalytic activity of the metal for reactions between those two gases. Beyond these general effects, we enquire whether any local modifications of catalytic activity can be ascribed to irregularities which modify our previous conclusions regarding penetrability of layers on a plane lattice.

As previously, the behaviour of front and rear portions of a long chain molecule, adhering perpendicularly to an interface, throws light on the behaviour of the two superposed components of a double layer. If the surface contains an isolated projection, the head of the chain can reach the mean plane while the tail lies alongside the raised excrescence, and similarly a fissure allows the head to go below the mean surface while the tail is level with the plane. In both cases the outer portion of the molecule has access to atoms of the solid from which it is otherwise withheld. Since decomposition of organic molecules forms the main experimental basis of the projecting-patch theory of catalysis (and indeed the applicability of that theory to reactions other than the decomposition in question has been challenged), the explanation may partly be attributable to such penetration by an outer portion of an orientated molecule which would, at a perfectly plane surface, never reach the metal.

Extending the term "penetration" from the behaviour of the rear portion of a long single molecule to that of the outer component of a double layer, let us take three conventionalized types of irregularity, of which any actual surface region will present instances similar in principle to but more complex in detail than the proposed models.

Projection above plane lattice. Consider an idealized lattice on which projects a single row of additional atoms. On the solid is adsorbed a saturated layer of larger atoms, and on this layer impinge smaller atoms. So long as the projecting atoms and the two gas components differ considerably in size, the former introduce gaps in the regular packing of the otherwise saturated layer, allowing the second component of smaller size to enter between the projection and the first component. Access to the solid will occur at temperatures too low for normal interpenetration of the two gases to be attained by the mechanism discussed in connexion with migration of the lower layer, and a number of the smaller atoms will accumulate in any such hollows of atomic size caused by the presence of the projection. Reactions in which these small atoms take part will tend to be accelerated along such linear projections, since this trap greatly extends the duration of their life in the unevaporated state. Such an effect will be greatest for the greatest discrepancy in size between the gaseous components, and for the greatest mobility of the smaller atom. Since the latter condition will be best fulfilled if the projecting row of atoms are not themselves too strongly attracting, the effect of the projection may be greater if it consists of partly inert foreign particles than if it consists of the underlying metal itself: thus we approach the conception of a non-reacting promoter facilitating the catalytic powers of a more active solid.

Parallel-sided fissure of atomic width. Consider a fissure whose width is not greater than that of the larger atoms of the first gas component. In this case the second component has not the additional mode of access which, in the case of a projection above a plane lattice, allowed it to undermine a saturated layer. But when once any ordinary penetration at the top of the fissure sets in, accumulation of the second and smaller component in the vacant space below can proceed, draining the gas phase into the metal by decreasing continuously the chance of re-evaporation of the dissolved gas. All consequences of penetration of the first layer by the second will thus be accelerated from such point sources or line sources.

Narrowing fissure. In most practical cases the open end of a fissure is wider than the largest atom: the principal difference between this case and the last is in the chance of escape of reactants and of products of reaction; this decides whether the reactions initiated at a fissure proceed indefinitely or are terminated. In general, if the fissure is wide enough for the larger atoms also to accumulate in it, their presence will accelerate reactions in which they take part but will retard reactions involving only the other components and the solid, by blocking the way to evaporation.

In addition to the above types of mechanism, by which surface irregularities anticipate and accelerate the plane penetrations discussed in previous sections, all fissures and projections cause local suspension of the inhibitions of interpenetration of layers by lateral repulsion.

By introducing third and fourth components of varied properties, the models can be extended, in principle, to allow discussion of catalytic poisons as well as of promoters. Another extension is indicated where lattice irregularities cause transmission and accumulation of material which was initially in the interior of the metal instead of in the gas phase. A molecular mechanism of penetrability is required for the phenomena of release from solid solution, e.g. the renewal of surface layers from within a thoriated filament, and the effect thereon of existing deposits from outside.

The very much simplified structures to which we have confined attention are sufficient to indicate that the common preparation of a solid for chemical activity by heating and by alternate oxidation and reduction must cover a very complex sequence of physical processes, which will rarely be either reversible or repeatable; a single treatment may succeed in mobilizing each component to take advantage of fissure growth, but its repetition at excess temperature may lead to inhibition of activity by sintering or closure of fissures. Up to a certain point, increase of subdivision of a solid facilitates most of the mechanisms we have described, to a limit based on the following considerations. Let A be the activity of a surface, defined by the velocity of reaction of given components striking unit area under given conditions of temperature and pressure. Then if unit activity is that of which the perfect lattice is capable and A' the added activity at grain-boundaries or fissure-edges,

$$A = 1 + A',$$

$$A' = f\left(\frac{n_1}{n_2}\right),$$

where n_1 is the number of particles striking an edge and n_2 the number striking the plane area, and f is some function.

In the simplest case where square grains are separated by cracks of the kind first studied by Smekal and by Zwicky, n_1/n_2 will increase a hundredfold during any subdivision of the surface which reduces average grain-length from 10^{-3} cm. to 10^{-5} cm., for atoms about 3×10^{-8} cm. in diameter. Where f is a constant, and made large by favourable development of some of the irregular penetrations which we have suggested, A will be increased by any treatment which tends to break up the solid, up to a limiting ratio of n_1/n_2 set by the size of the gas particles.

§ 6. CONCLUSION

It has been shown that only in certain restricted circumstances can two substances adsorbed from the gaseous phase maintain stable existence as superposed strata, even if the one is completely condensed before the other is admitted to the surface. In general the two substances will interpenetrate with greater or lesser rapidity, in certain cases sharing a single layer of monomolecular thickness, and in others reversing their order of proximity to the underlying solid. The occurrence of these interpenetrations has been related to regular and irregular structure of the solid, to dimensions and force laws of any two adsorbed components, and to their mobilities and the several inhibitions controlling the latter. All the results have been obtained by very much simplified models and conventionalized force laws; only to the extent that actual oxygen, hydrogen, inert gases, alkali vapours, etc., have properties describable in terms of these simple models, will the conclusions apply to these substances.

It is hoped next to deal quantitatively with certain of the problems treated here only in general outline.

REFERENCES

(1) General reference, Reimann, *Thermionic Emission* (1934).
(2) General reference, Hughes and DuBridge, *Photoelectric Phenomena* (1932).
(3) General reference, *Faraday Society Discussion on Gas Adsorption* (1932).
(4) Slater. *Phys. Rev.* **36**, 57 (1930).
(5) Langmuir and Villars. *J. Amer. chem. Soc.* **53**, 486 (1931).
(6) Taylor and Langmuir. *Phys. Rev.* **44**, 423 (1933).
(7) Lennard-Jones. *Trans. Faraday Soc.* **28**, 333 (1932).

535·52

CONVERGENT POLARIZED LIGHT AND HERTZ'S PROBLEM FOR A UNIAXIAL MATERIAL

By F. F. P. BISACRE, O.B.E., M.A., B.Sc.

Received November 1, 1934. *Read in title December* 21, 1934.

ABSTRACT. The paper gives a mathematical treatment of the behaviour of convergent polarized light passing through a uniaxial crystal, together with some practical details relating to the production of rings and brushes. The Fresnel wave surface is shown to be the isophasic for a system of waves diverging from a point source in the crystal. Explicit expressions are given, in terms of elementary functions, for the electric and magnetic fields and the energy-stream, both far from and close to the source; and the validity of Fresnel's construction is discussed.

§ 1. INTRODUCTION

On account of purely technical mathematical difficulties the classical theory of the optics of crystals is usually based almost entirely upon plane waves; the Fresnel wave surface is deduced as the envelope of a system of plane waves passing through the crystal in all directions.

In making instruments to exhibit the well-known rings and brushes, opticians have carefully reproduced the conditions that have had to be assumed in the mathematical theory. But the usual rings and brushes can be produced as easily and as brilliantly with truly divergent polarized light as with the artificial system of plane waves that is usually described as "convergent polarized light," much to the mystification of the student.

An attempt is made in this paper to show that in the case of a uniaxial crystal the Fresnel wave surface is not merely the envelope of a system of plane waves passing through the material, but the actual isophasic for a system of waves diverging from a point-source in the crystal. No attempt is made here to deal with the much more difficult mathematical problem of a biaxial crystal*. The problem is, of course, to solve Hertz's problem for a uniaxial material.

---

* One of the earliest treatments of this problem (from the elastic solid standpoint) was given, nearly a century ago, by Lamé, and can be found in his treatise *Leçons sur la Théorie mathématique de l'Élasticité des Corps solides* (Paris, 1852, 2nd ed., 1866). Certain general solutions have been discussed by Herglotz, who used mathematical methods of extreme difficulty. His papers are: "Ueber die Integration linearer partieller Differentialgleichungen mit Konstanten Koefficienten, Teile I, II, III." *Ber. sächs. Ges.* (*Akad.*) *Wiss.* **78** (1926) (I, II), **80** (1928) (III); "Ueber die Integration linearer partieller Differential-gleichungen mit Konstanten Koefficienten," *Abh. math. Sem. hamburg. Univ.* **6**, Heft 7 (1928). See also S. Kowalewski, "Über die Brechung des Lichtes in kristallinischen Mitteln", *Acta Math.* (1885); V. Volterra, "Sur les vibrations dans les milieux biréfringents", *Acta Math.* (1892); J. Grünwald, "Ausbreitung der Wellenbewegungen in optisch zweiachsigen elastichen Medien", *Boltzmann-Festschrift* (Leipzig, 1904).

§2. CALCULATION OF THE MAGNETIC AND ELECTRIC FIELDS

We begin with Lorentz's equations for a set of principal axes,

H, E

$$\left.\begin{aligned} c\left(\frac{\partial H_z}{\partial y}-\frac{\partial H_y}{\partial z}\right)&=4\pi\rho v_x+K_1\dot{E}_x\\ c\left(\frac{\partial H_x}{\partial z}-\frac{\partial H_z}{\partial x}\right)&=4\pi\rho v_y+K_1\dot{E}_y\\ c\left(\frac{\partial H_y}{\partial x}-\frac{\partial H_x}{\partial y}\right)&=4\pi\rho v_z+K_3E_z \end{aligned}\right\} \quad \text{......(1),}$$

c, ρ
v, K
x, y, z

so that Oz is the axis of symmetry of the material.

$$-c\operatorname{curl}\mathbf{E}=\dot{\mathbf{H}} \quad \text{......(2),}$$

$$\operatorname{div}\mathbf{H}=0 \quad \text{......(3),}$$

$$\operatorname{div}\mathbf{D}=\rho \quad \text{......(4),}$$

where

$$D_x=K_1E_x/4\pi,\ \text{etc.} \quad \text{......(5),}$$

D

and by (4)

$$K_1\frac{\partial E_x}{\partial x}+K_1\frac{\partial E_y}{\partial y}+K_3\frac{\partial E_z}{\partial z}=4\pi\rho \quad \text{......(6).}$$

We assume that ρ and $\mathbf{v}$ are given as functions of position and time, where $\mathbf{v}$ is the velocity of electric charge. Assume, as usual, that $\mathbf{A}$ is a vector potential such that

v
A

$$\mathbf{H}=\operatorname{curl}\mathbf{A},$$

then

$$\dot{\mathbf{H}}=\operatorname{curl}\dot{\mathbf{A}},$$

and, by (2),

$$-c\operatorname{curl}\mathbf{E}=\operatorname{curl}\dot{\mathbf{A}},$$

i.e.

$$\operatorname{curl}(\dot{\mathbf{A}}+c\mathbf{E})=0,$$

or, since curl grad $\phi=0$,

$$\mathbf{E}=-\dot{\mathbf{A}}/c-\operatorname{grad}\phi \quad \text{......(7),}$$

φ

where ϕ is an arbitrary scalar function of x, y, z and t.

Thus (1) becomes

$$\left.\begin{aligned} c\left(\frac{\partial H_z}{\partial y}-\frac{\partial H_y}{\partial z}\right)&=4\pi\rho v_x+K_1\left(-\frac{\ddot{A}_x}{c}-\frac{\partial\phi}{\partial x}\right)\\ c\left(\frac{\partial H_x}{\partial z}-\frac{\partial H_z}{\partial x}\right)&=4\pi\rho v_y+K_1\left(-\frac{\ddot{A}_y}{c}-\frac{\partial\phi}{\partial y}\right)\\ c\left(\frac{\partial H_y}{\partial x}-\frac{\partial H_x}{\partial y}\right)&=4\pi\rho v_z+K_3\left(-\frac{\ddot{A}_z}{c}-\frac{\partial\phi}{\partial z}\right) \end{aligned}\right\} \quad \text{......(8),}$$

or, since

$$\begin{aligned}\operatorname{curl}\mathbf{H}&=\operatorname{curl}\operatorname{curl}\mathbf{A}\\ &=\operatorname{grad}\operatorname{div}\mathbf{A}-\Delta\mathbf{A},\end{aligned}$$

Δ

$$\left.\begin{aligned} 4\pi\rho v_x-\frac{K_1}{c}\ddot{A}_x+c\Delta A_x&=\frac{\partial}{\partial x}(c\operatorname{div}\mathbf{A}+K_1\phi)\\ 4\pi\rho v_y-\frac{K_1}{c}\ddot{A}_y+c\Delta A_y&=\frac{\partial}{\partial y}(c\operatorname{div}\mathbf{A}+K_1\phi)\\ 4\pi\rho v_z-\frac{K_3}{c}\ddot{A}_z+c\Delta A_z&=\frac{\partial}{\partial z}(c\operatorname{div}\mathbf{A}+K_3\phi) \end{aligned}\right\} \quad \text{......(9).}$$

A has, so far, been incompletely defined; its curl only is specified so far. We complete its specification by defining its divergence and this we do by putting

$$c \operatorname{div} \mathbf{A} + K_1 \phi = 0 \qquad \text{......(10)},$$

hence the first two equations of (9) become

$$\left.\begin{aligned} c\Delta A_x - \frac{K_1}{c}\ddot{A}_x &= -4\pi\rho v_x \\ c\Delta A_y - \frac{K_1}{c}\ddot{A}_y &= -4\pi\rho v_y \end{aligned}\right\} \qquad \text{......(11)},$$

and the third $$c\Delta A_z - \frac{K_3}{c}\ddot{A}_z = -4\pi\rho v_z + c\,\frac{K_1 - K_3}{K_1}\frac{\partial}{\partial z}\operatorname{div}\mathbf{A} \qquad \text{......(12)},$$

or, somewhat more conveniently,

$$\left.\begin{aligned} \left(-\frac{K_1}{c^2}\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) A_x &= -\frac{4\pi\rho v_x}{c} & (a) \\ \left\{-\frac{K_1}{c^2}\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right\} A_y &= -\frac{4\pi\rho v_y}{c} & (b) \\ \left[\left\{-\frac{K_1}{c^2}\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial z^2}\right\} + \frac{K_1}{K_3}\left\{\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right\}\right] A_z &= -\frac{4\pi\rho v_z}{c} \times \frac{K_1}{K_3} \\ &\quad + \frac{K_1 - K_3}{K_3}\frac{\partial}{\partial z}\left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y}\right) & (c) \end{aligned}\right\} \quad \text{...(13)}.$$

We shall suppose a Hertzian electric dipole(1) at the origin with its axis pointing in any direction. We can choose the axes so that $v_y = 0$, i.e. so that the axis of the dipole lies in the xz plane. Equation 13 (*b*) now disappears as well as the last term in the last bracket on the right of equation 13 (*c*). Equation 13 (*a*) in A_x is a standard wave equation and has a well-known solution having a suitable singularity at the origin and vanishing at infinity to the proper order. It is

$$A_x = \frac{\dot{M}_x}{cr} e^{ip(t - r/a)} \qquad \text{......(14)},$$

$\dot{M}$, M where by $\dot{M}_x$ we mean ipM_x, M_x being the maximum value of the x component of the electric moment of the dipole, and

$$\left.\begin{aligned} \frac{c^2}{K_1} &= a^2 \\ r^2 &= x^2 + y^2 + z^2 \end{aligned}\right\} \qquad \text{......(14}a\text{)}.$$

r and

We are left with equation 13 (*c*) to solve. Since a solution having the required singularity at the origin and vanishing at infinity in the proper way is unique, we only need a particular integral of equation 13 (*c*). We can obtain it in the following
$A_{z,0}$ way. A_z can be considered as made up of two parts, $A_{z,0}$ coming from the *first*
$A_{z,1}$ term on the right of equation 13 (*c*) and $A_{z,1}$ from the second term. We have then

$$\left[-\frac{1}{a^2}\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial z^2} + \frac{K_1}{K_3}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\right] A_{z,0} = -\frac{K_1}{K_3}\frac{4\pi\rho v_z}{c},$$

or, putting

$$x, y = \sqrt{\left(\frac{K_1}{K_3}\xi\right)},\ \sqrt{\left(\frac{K_1}{K_3}\eta\right)},$$

ξ, η

$$= \frac{1}{\beta}\xi,\ \frac{1}{\beta}\eta,$$

β

we have

$$\left(\Delta' - \frac{1}{a^2}\frac{\partial^2}{\partial t^2}\right) A_{z,0} = -\frac{K_1}{K_3}\frac{4\pi\rho v_z}{c} \qquad \text{......(15)},$$

where

$$\Delta' \equiv \frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2} + \frac{\partial^2}{\partial z^2}.$$

This is a standard case like the last, so

$$A_{z,0} = \frac{1}{\beta^2}\frac{\dot{M}_z}{cP} e^{ip(t-P/a)} \qquad \text{......(16)},$$

where

$$P^2 = \xi^2 + \eta^2 + z^2$$
$$= \beta^2(x^2+y^2) + z^2.$$

P

This wave is a point-source spherical wave in ξ, η, z space and so a spheroidal wave in x, y, z space. We are now left to find $A_{z,1}$, where

$$\left[-\frac{1}{a^2}\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial z^2} + \frac{1}{\beta^2}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\right] A_{z,1} = \left(\frac{1-\beta^2}{\beta^2}\right)\frac{\partial^2}{\partial z\,\partial x}\left\{\frac{\dot{M}_x}{cr} e^{ip(t-r/a)}\right\} \quad \text{...(17)}.$$

Put

$$\left.\begin{aligned} \frac{1-\beta^2}{\beta^2}\frac{\partial B}{\partial z} &= A_{z,1} \\ B' &= Bc/\dot{M}_x \\ &= \partial g/\partial x \end{aligned}\right\} \qquad \text{......(18)},$$

and

B

B'

g

then

$$\left[-\frac{1}{a^2}\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial z^2} + \frac{1}{\beta^2}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\right] g = \frac{1}{r} e^{ip(t-r/a)}$$

or

$$\left[-\frac{1}{a^2}\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial z^2} + \frac{1}{\beta^2}\left(\frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial}{\partial \rho}\right)\right] g = \frac{1}{\sqrt{(\rho^2+z^2)}} e^{ip(t-(\rho^2+z^2)^{\frac{1}{2}}/a)} \quad \text{...(19)},$$

where

$$r^2 = x^2 + y^2 + z^2 = \rho^2 + z^2.$$

ρ

Now it is possible that, if

$$\left(\frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial}{\partial \rho}\right) g' = \frac{1}{\sqrt{(\rho^2+z^2)}} e^{ip(t-(\rho^2+z^2)^{\frac{1}{2}}/a)},$$

then

$$\left(-\frac{1}{a^2}\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial z^2}\right) g' = \frac{G}{\sqrt{(\rho^2+z^2)}} e^{ip(t-(\rho^2+z^2)^{\frac{1}{2}}/a)} \qquad \text{......(19 } a\text{)},$$

where G is some *constant**; and this supposition proves to be true, as we shall now show. We can easily solve

$$\rho\frac{\partial^2 g'}{\partial \rho^2} + \frac{\partial g'}{\partial \rho} = \frac{\rho}{\sqrt{(\rho^2+z^2)}} e^{ip(t-(\rho^2+z^2)^{\frac{1}{2}}/a)},$$

i.e.

$$\frac{\partial}{\partial \rho}\left(\rho\frac{\partial g'}{\partial \rho}\right) = \frac{\partial}{\partial \rho}\left\{-\frac{a}{ip} e^{ip(t-(\rho^2+z^2)^{\frac{1}{2}}/a)}\right\},$$

i.e.

$$\rho\frac{\partial g'}{\partial \rho} = -\frac{a}{ip} e^{ip(t-(\rho^2+z^2)^{\frac{1}{2}}/a)},$$

i.e.

$$g' = -\frac{a}{ip}\int_{\infty}^{\rho}\frac{1}{\rho} e^{ip(t-(\rho^2+z^2)^{\frac{1}{2}}/a)}\, d\rho \qquad \text{......(20)}.$$

---

* Dr John Dougall suggested this to me.

Differentiating twice with respect to z, we get

$$\frac{\partial^2 g'}{\partial z^2}=\int_\infty^\rho \frac{\rho}{r^3}\, e^{ip\,(t-r/a)}\, d\rho-\frac{ip}{a}\int_\infty^\rho \frac{z^2}{\rho r^2}\, e^{ip\,(t-r/a)}\, d\rho \qquad \text{......(21)},$$

and differentiating twice with respect to t and dividing by $-a^2$ we get

$$-\frac{1}{a^2}\frac{\partial^2 g'}{\partial t^2}=\frac{ip}{a}\int_\infty^\rho \frac{1}{\rho}\, e^{ip\,(t-r/a)}\, d\rho \qquad \text{......(22)}.$$

Adding equations (21) and (22), we get

$$\left(-\frac{1}{a^2}\frac{\partial^2}{\partial t^2}+\frac{\partial^2}{\partial z^2}\right) g'=\int_\infty^\rho \left(\frac{\rho}{r^3}+\frac{ip}{a}\cdot\frac{\rho}{r^2}\right) e^{ip\,(t-r/a)}\, d\rho$$

$$=\int_\infty^\rho \frac{\partial}{\partial \rho}\left\{-\frac{1}{r}\, e^{ip\,(t-r/a)}\right\} d\rho$$

$$=-\frac{1}{r}\, e^{ip\,(t-r/a)};$$

hence G of equation 19 (a) is given by

$$G=-1 \qquad \text{......(23)},$$

so the hypothesis is proved.

We can now easily get a solution of equation (19).

$\lambda,\ g'$ Let $g=\lambda g'$, where g' is the function given by equation (20); then, substituting in equation (19), we get

$$\left[-\lambda\left(\frac{1}{r}\, e^{ip\,(t-r/a)}\right)+\frac{\lambda}{\beta^2}\left(\frac{1}{r}\, e^{ip\,(t-r/a)}\right)\right]=\frac{1}{r}\, e^{ip\,(t-r/a)},$$

i.e.

$$\lambda\left(\frac{1}{\beta^2}-1\right)=1,$$

$$\therefore\ \lambda=\frac{\beta^2}{1-\beta^2} \qquad \text{......(24)},$$

and

$$g=\frac{\beta^2}{1-\beta^2}\, g'$$

solves equation (19), where

$$g'=-\frac{a}{ip}\int_\infty^\rho \frac{1}{\rho}\, e^{ip\,(t-\sqrt{(\rho^2+z^2)}/a)}\, d\rho \qquad \text{......(25)},$$

and, by equation (18),

$$B'=\frac{\beta^2}{1-\beta^2}\cdot\frac{\partial g'}{\partial x}=\frac{Bc}{\dot{M}_x} \qquad \text{......(26)};$$

$$\therefore\ B=\frac{\dot{M}_x}{c}\cdot\frac{\beta^2}{1-\beta^2}\cdot\frac{\partial g'}{\partial x},$$

$A_{z,1}$ and

$$A_{z,1}=\frac{\dot{M}_x}{c}\cdot\frac{\partial^2 g'}{\partial z\,\partial x}=\frac{\dot{M}_x}{c}\cdot\frac{\partial}{\partial z}\cdot\left(\frac{\partial g'}{\partial \rho}\cdot\frac{\partial \rho}{\partial x}\right)$$

$$=\frac{\dot{M}_x}{c}\cdot\frac{xz}{\rho^2}\cdot\frac{1}{r}\cdot e^{ip\,(t-r/a)} \qquad \text{......(27)}.$$

This function has a singularity along the whole z-axis. We can remove it in the following way:

Since $\xi=\beta x$, $\eta=\beta y$, we get

$$\left[-\frac{1}{a^2}\frac{\partial^2}{\partial t^2}+\frac{\partial^2}{\partial z^2}+\frac{1}{\beta^2}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\right]u\equiv\left(-\frac{1}{a^2}\frac{\partial^2}{\partial t^2}+\frac{\partial^2}{\partial z^2}+\frac{\partial^2}{\partial \xi^2}+\frac{\partial^2}{\partial \eta^2}\right)u\,(\xi,\eta,z).$$

We know from equations (20) and (23) that

$$u=-\frac{a}{ip}\int_\infty^R\frac{1}{R}\,e^{ip(t-P/a)}\,dR \qquad \text{......(28)},$$

where $R^2=\xi^2+\eta^2=\beta^2(x^2+y^2)$

and $P^2=R^2+z^2$ ......(29)

lead to the equation $$\left(\frac{\partial^2}{\partial \xi^2}+\frac{\partial^2}{\partial \eta^2}\right)u=-\left(-\frac{1}{a^2}\frac{\partial^2}{\partial t^2}+\frac{\partial^2}{\partial z^2}\right)u,$$

i.e. $$\left(-\frac{1}{a^2}\frac{\partial^2}{\partial t^2}+\Delta'\right)u=0,$$

i.e. $$\left[-\frac{1}{a^2}\frac{\partial^2}{\partial t^2}+\frac{\partial^2}{\partial z^2}+\frac{1}{\beta^2}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\right]u=0.$$

We can consequently add any multiple we please of

$$\frac{\dot M_x}{c}\cdot\frac{\xi z}{R^2}\cdot\frac{1}{P}\,e^{ip(t-P/a)}$$

to the solution already obtained; compare equation (27). Let us add $-\beta$ times this expression, where

$$-\beta\xi z/R^2=-xz/\rho^2.$$

The complete solution is then

$$A_{z,1}=\frac{\dot M_x}{c}\cdot\frac{xz}{\rho^2}\left(\frac{1}{r}\,e^{ip(t-r/a)}-\frac{1}{P}\,e^{ip(t-P/a)}\right) \qquad \text{......(30)},$$

which has a singularity at the origin only. Our vector potential is then

$$A_x=\frac{\dot M_x}{cr}\cdot e^{ip(t-r/a)},$$

$$A_z=\frac{1}{\beta^2}\cdot\frac{\dot M_z}{cP}\;{}^{i\gamma(t-P/a)}$$

$$+\frac{\dot M_x}{c}\cdot\frac{xz}{\rho^2}\left(\frac{1}{r}\,e^{ip(t-r/a)}-\frac{1}{P}\,e^{ip(t-P/a)}\right) \qquad (31),$$

where $\beta^2=K_3/K_1$

and $P^2=\beta^2(x^2+y^2)+z^2.$

The electromagnetic field is at once given from this vector potential, since

$\mathbf{H}=\text{curl}\,\mathbf{A}$

and $\mathbf{E}=-\dot{\mathbf{A}}/c-\text{grad}\,\phi$ ......(32).

where $c\,\text{div}\,\mathbf{A}+K_1\dot\phi=0$

The first of these equations gives the readiest means of calculating **H** from **A**; but, once **H** is known, **E**, for points at a distance, can be more simply found directly from Lorentz's equations observing that the first three become

$$c \operatorname{curl} \mathbf{H} = (K_1, K_1, K_3)\,\dot{\mathbf{E}}$$

except at the origin, giving $\dot{\mathbf{E}}$ and hence **E** immediately. Since A_x alone involves purely spherical waves, it is immediately clear that H_z, alone of all the six components of the field, proceeds exclusively on a spherical isophasic. We propose to examine the fields due to these vector potentials, at a considerable distance from the origin, bearing in mind that for this purpose we need only differentiate the exponential functions, treating the rest of the expressions for the A's as constant.

§ 3. THE SPHERICAL WAVE

F Put $A_x' = F.e^{-i\alpha r}.e^{ipt}$, where $F.e^{ipt} = \dfrac{\dot{M}_x}{cr}\,e^{ipt}$,

G $A_z' = G.e^{-i\alpha r}.e^{ipt}$, where $G.e^{ipt} = \dfrac{xz}{\rho^2}.\dfrac{\dot{M}_x}{cr}e^{ipt}$

α and $\alpha = p/a$.

Then, the common factor e^{ipt} being omitted,

$$\left.\begin{aligned} H_x &= (-i\alpha).\frac{xyz}{r\rho^2}.Fe^{-i\alpha r} \\ H_y &= (-i\alpha).\frac{y^2z}{r\rho^2}.Fe^{-i\alpha r} \\ H_z &= (-i\alpha).\left\{-\frac{v\,(x^2+y^2)}{r\rho^2}\right\}.Fe^{-i\alpha r} \end{aligned}\right\} \quad \ldots\ldots(33).$$

This result shows that, if $\mathbf{R} = (x, y, z)$,

$$(\mathbf{RH}) = 0,$$

i.e. **H** is perpendicular to the radius vector, **R**.

From equation (33) we get at once the real forms of H:

$$\left.\begin{aligned} H_x &= \frac{p^2}{a}.\frac{M_x}{cr}.\frac{xyz}{\rho^2 r}.\cos p\left(t-\frac{r}{a}\right) \\ H_y &= \frac{p^2}{a}.\frac{M_x}{cr}.\frac{y^2z}{\rho^2 r}.\cos p\left(t-\frac{r}{a}\right) \\ H_z &= -\frac{p^2}{a}.\frac{M_x}{cr}.\frac{y\,(x^2+y^2)}{\rho^2 r}.\cos p\left(t-\frac{r}{a}\right) \end{aligned}\right\} \quad \ldots\ldots(34).$$

The amplitude of H is therefore

$$|H| = \frac{p^2}{a}.\frac{y}{\rho}.\frac{M_x}{cr} \quad \ldots\ldots(35)$$

and

$$h = \frac{p^2}{a}.\frac{y}{\rho}.\frac{M_x}{cr}\cos p\left(t-\frac{r}{a}\right) \quad \ldots\ldots(35a),$$

where h is the numerical value of the whole magnetic vector at any time and place.

To calculate the electric vector, we have

$$c \operatorname{curl} \mathbf{H} = (K_1,\ K_1,\ K_3)\, ip\mathbf{E},$$

i.e.

$$\left.\begin{aligned} E_x &= \frac{c}{ipK_1}\,(\operatorname{curl}\mathbf{H})_x \\ E_y &= \frac{c}{ipK_1}\,(\operatorname{curl}\mathbf{H})_y \\ E_z &= \frac{c}{ipK_3}\,(\operatorname{curl}\mathbf{H})_z \end{aligned}\right\} \qquad \ldots\ldots(36),$$

whence

$$\left.\begin{aligned} E_x &= -\frac{ip}{c}\cdot\frac{y^2r^2}{\rho^2r^2}\cdot Fe^{-i\alpha r} \\ E_y &= +\frac{ip}{c}\cdot\frac{xyr^2}{\rho^2r^2}\cdot Fe^{-i\alpha r} \\ E_z &= 0 \end{aligned}\right\} \qquad \ldots\ldots(37);$$

so the real forms are:

$$\left.\begin{aligned} E_x &= \frac{p^2}{c}\cdot\frac{M_x}{cr}\cdot\frac{y^2r^2}{\rho^2r^2}\cdot\cos p\left(t-\frac{r}{a}\right) \\ E_y &= -\frac{p^2}{c}\cdot\frac{M_x}{cr}\cdot\frac{xyr^2}{\rho^2r^2}\cos p\left(t-\frac{r}{a}\right) \\ E_z &= 0 \end{aligned}\right\} \qquad \ldots\ldots(38).$$

The scalar product (**RE**) is again zero, so **E** is perpendicular to the radius vector. Also (**EH**) = 0; therefore **E** and **H** are perpendicular. Also, since $E_z = 0$, **E** lies wholly in the plane parallel to xy, i.e. **E** lies along the parallel of latitude and **H** lies along the meridian through the point. The amplitude of **E** is

$$\frac{p^2}{c}\cdot\frac{M_x}{cr}\cdot\frac{y}{\rho} \qquad \ldots\ldots(39)$$

and

$$e = \frac{p^2}{c}\cdot\frac{M_x}{cr}\cdot\frac{y}{\rho}\cdot\cos p\left(t-\frac{r}{a}\right) \qquad \ldots\ldots(40),$$

where e is the numerical value of the whole electric vector. The vector product is e

$$|[\mathbf{EH}]| = \mathbf{eh} = \frac{p^4}{ca}\cdot\left(\frac{M_x}{cr}\right)^2\cdot\frac{y^2}{\rho^2}\cdot\cos^2 p\left(t-\frac{r}{a}\right) \qquad \ldots\ldots(41)$$

and, if θ, ϕ

$$\begin{aligned} x &= r\sin\theta\cos\phi, \\ y &= r\sin\theta\sin\phi, \\ z &= r\cos\theta, \\ y/\rho &= \sin\phi; \end{aligned}$$

hence, since the Poynting vector **S** is given by **S**

$$\mathbf{S} = \frac{c}{4\pi}\,[\mathbf{EH}],$$

energy radiates outwards on the spherical wave at a mean rate $S_{m,1}$, where

$S_{m,1}$

$$S_{m,1}=\frac{c}{8\pi}\cdot\frac{p^4}{ca}\cdot\left(\frac{M_x}{cr}\right)^2\sin^2\phi$$
$$=\frac{1}{8\pi}\cdot\frac{p^4}{c^2a}\cdot\frac{M_x^2}{r^2}\cdot\sin^2\phi \qquad \text{......(42).}$$

It is evident that $S_{m,1}$ is constant all over any meridian; on a given isophasic it varies with the longitude only.

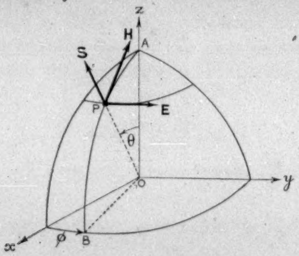


Figure 1. Spherical field.

§4. THE SPHEROIDAL WAVE

We now want a similar calculation for the spheroidal wave.

Put
$$A_z''=\left(\frac{1}{\beta^2}\cdot\frac{\dot{M}_z}{cP}-\frac{xz}{\rho^2}\cdot\frac{\dot{M}_x}{cP}\right)e^{-i\alpha P}\cdot e^{ipt} \qquad \text{......(43)}$$
$$=ipBe^{-i\alpha P}\cdot e^{ipt},$$

where
$$Be^{ipt}=\left(\frac{1}{\beta^2}\cdot\frac{M_z}{cP}-\frac{xz}{\rho^2}\cdot\frac{M_x}{cP}\right)e^{ipt} \qquad \text{......(44),}$$

and observe that
$$\frac{\partial P}{\partial x}=\beta^2\frac{x}{P},\quad \text{where}\quad \beta^2=K_3/K_1,$$

β

$$\frac{\partial P}{\partial y}=\beta^2\frac{y}{P},$$
$$\frac{\partial P}{\partial z}=\frac{z}{P},$$

since
$$P^2=\beta^2(x^2+y^2)+z^2.$$

Then, the common factor e^{ipt} being omitted,

$$\left.\begin{aligned}H_x&=\frac{\partial A_z}{\partial y}=(-i\alpha)\cdot\beta^2\cdot\frac{y}{P}\cdot ipBe^{-i\alpha P}\\ H_y&=-\frac{\partial A_z}{\partial x}=-(-i\alpha)\,\beta^2\frac{x}{P}\cdot ipBe^{-i\alpha P}\\ H_z&=0\end{aligned}\right\} \qquad \text{......(45).}$$

This shows that $(\mathbf{RH})=0$, i.e. $\mathbf{H}$ is perpendicular to the radius vector, $\mathbf{R}$.

The real form of **H** is

$$H_x=\frac{p^2}{a}.\beta^2.\frac{y}{P}.B.\cos p\left(t-\frac{P}{a}\right),$$

$$H_y=-\frac{p^2}{a}.\beta^2.\frac{x}{P}.B.\cos p\left(t-\frac{P}{a}\right),$$

$$H_z=0.$$

The amplitude of **H** is $$\frac{p^2}{a}.\beta^2.\frac{\rho}{P}.B,$$

so that $$h=\frac{p^2}{a}.\beta^2.\frac{\rho}{P}.B.\cos p\left(t-\frac{P}{a}\right) \qquad \text{......(46)},$$

where B is defined by equation (44).

We can calculate the electric vector by the formula (36).

$$\left.\begin{aligned} E_x&=+\frac{ip}{c}.\beta^2.\frac{xz}{P^2}.ipB.e^{-i\alpha P}\\ E_y&=+\frac{ip}{c}.\beta^2.\frac{yz}{P^2}.ipB.e^{-i\alpha P}\\ E_z&=-\frac{ip}{c}.\beta^2.\frac{\rho^2}{P^2}.ipB.e^{-i\alpha P}\end{aligned}\right\} \qquad \text{......(47)}.$$

Consequently the real values are:

$$\left.\begin{aligned} E_x&=-\frac{p^2}{c}.\beta^2.\frac{xz}{P^2}.B.\cos p\left(t-\frac{P}{a}\right)\\ E_y&=-\frac{p^2}{c}.\beta^2.\frac{yz}{P^2}.B.\cos p\left(t-\frac{P}{a}\right)\\ E_z&=+\frac{p^2}{c}.\beta^2.\frac{\rho^2}{P^2}.B.\cos p\left(t-\frac{P}{a}\right)\end{aligned}\right\} \qquad \text{......(48)}.$$

The amplitude of **E** is $$\frac{p^2}{c}.\beta^2.\frac{\rho r}{P^2}.B;$$

$$\therefore\ e=\frac{p^2}{c}.\beta^2.\frac{\rho r}{P^2}.B\cos p\left(t-\frac{P}{a}\right).$$

Equation (47) shows that **E** is perpendicular to **R**; also, by equations (45) and (47),

$$(\mathbf{EH})=0,$$

so that **E** is perpendicular to **H**.

The vector product of **E** and **H** is

$$\frac{p^4}{ac}.\beta^4.\frac{\rho^2 r}{P^3}.B^2.\cos^2 p\left(t-\frac{P}{a}\right)$$

and consequently the mean rate, $S_{m,2}$, of radiation of energy is $S_{m,2}$

$$S_{m,2}=\frac{1}{8\pi}.\frac{p^4}{c^2aP^2}.\left(\frac{r}{P}\right)^3\sin^2\theta\,\{M_z-\beta^2M_x\cos\phi\cot\theta\}^2 \qquad \text{......(49)},$$

where $$P^2=\beta^2(x^2+y^2)+z^2 \qquad \text{......(50)}.$$

Since $H_z = 0$, the **H** vector lies along a parallel of latitude of the spheroid and **E** in the meridian plane, perpendicular to **R**. **D**, however, is along the meridian, as can be easily shown, so that **D** and **H** lie in the tangent plane at P.

The Poynting vectors of each system are radial. Since the **E** of one field is parallel to the **H** of the other, the radiation of energy is given by the sum of $S_{m,1}$ due to the spherical field and $S_{m,2}$ due to the spheroidal field. This follows from the fact that

$$[\mathbf{A}+\mathbf{B},\ \mathbf{C}+\mathbf{D}] = [\mathbf{AC}] + [\mathbf{BD}]$$

if **B**, **C** are parallel and **A**, **D** are parallel. It is also clear that if we draw the tangent plane to the spheroid at P, the projection of the radius vector from O on to this tangent plane lies in the plane containing the meridian through P. This projection therefore gives the plane in which **E** and **D** lie, as it should by Fresnel's rule. It will be seen from equation (31) that if we make M_x equal to 0 the spherical field vanishes entirely, leaving only the spheroidal field due to M_z. If we now let $K_3 \to K_1$ (the isotropic case), this spheroidal field becomes spherical. Thus the spherical field of the isotropic case of an electric dipole is the degenerate spheroidal field—it is not the spherical field of the electric dipole in an aelotropic medium. Hence the **E** and **H** vectors are directed in accordance with the well-known results for the isotropic case. The directions of **E** and **H** for the aelotropic spherical field are perpendicular to these directions. If we start with an isotropic material and a dipole along Oz a spherical wave is generated. If we now introduce aelotropy by letting K_3 differ from K_1 while $K_1 = K_2$, this spherical wave becomes spheroidal. If we now let the dipole become oblique so that M_x exists, then a new spherical wave makes its appearance.

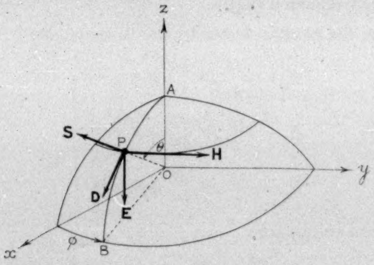


Figure 2. Spheroidal field. **S** is along OP and perpendicular to **E** and **H**; **D** and **H** lie in tangent plane at P; **D** and **E** lie in plane OAB.

§ 5. FIELD VERY CLOSE TO THE ORIGIN

Suppose $r \ll \lambda$, say $r = \lambda/100$. The most complicated form we have to differentiate, spatially, is

$$\frac{xz}{\rho^2}\cdot\frac{1}{r}\cdot e^{-i\alpha r} = \frac{xz}{\rho^2}\cdot\frac{1}{r}\left(1 - i\alpha r - \frac{\alpha^2 r^2}{2}\ldots\right)$$

$$= \frac{xz}{\rho^2}\cdot\frac{1}{r} - i\alpha\,\frac{xz}{\rho^2} - \text{positive powers of } r,$$

so that, when r and $\rho \to 0$ every quantity that matters comes from the first term; i.e. the retardation of phase becomes insignificant. It is therefore of little useful purpose to distinguish the two wave fields; it is more convenient to consider separately the field due to M_x and that due to M_z.

Field due to M_x. The vector potential is now given by

$$\left.\begin{aligned} A_x &= \frac{1}{r} \\ A_z &= \frac{xz}{\rho^2}\left(\frac{1}{r}-\frac{1}{P}\right)=\frac{(\beta^2-1)\,xz}{Pr\,(P+r)} \end{aligned}\right\} \qquad \ldots\ldots(51),$$

each multiplied by $ipM_x/c\,.\,e^{ipt}$.

These functions have singularities at the origin only.

To calculate $\mathbf{E}$ it is best to use the formula

$$\left.\begin{aligned} \mathbf{E} &= -\frac{\dot{\mathbf{A}}}{c}-\operatorname{grad}\phi \\ \text{where}\qquad \phi &= -\frac{c}{K_1}\operatorname{div}\mathbf{A} \end{aligned}\right\} \qquad \ldots\ldots(52).$$

Here grad ϕ is the important term, for it is two orders lower in r than $\dot{\mathbf{A}}$. Hence

$$\mathbf{E} \approx -\operatorname{grad}\phi,$$

where

$$\phi = -\frac{c}{K_1}\operatorname{div}\mathbf{A},$$

and

$$\operatorname{div}\mathbf{A} = \frac{\partial A_x}{\partial x}+\frac{\partial A_z}{\partial z} = -\beta^2 . \frac{x}{P^3} \times \frac{ipM_x}{c} . e^{ipt}.$$

If we put

$$\phi = ip\Phi e^{ipt}$$

we get

$$\Phi = \frac{K_3}{K_1^2} . M_x . \frac{x}{P^3} \qquad \ldots\ldots(53),$$

and

$$\left.\begin{aligned} E_x &= -\frac{K_3}{K_1^2} . M_x . \frac{\partial}{\partial x}\left(\frac{x}{P^3}\right) e^{ipt} \\ E_y &= -\frac{K_3}{K_1^2} . M_x . \frac{\partial}{\partial y}\left(\frac{x}{P^3}\right) e^{ipt} \\ E_z &= -\frac{K_3}{K_1^2} . M_x . \frac{\partial}{\partial z}\left(\frac{x}{P^3}\right) e^{ipt} \end{aligned}\right\} \qquad \ldots\ldots(54).$$

When $K_3 = K_1$ equation (53) reduces to

$$\Phi = \frac{1}{K_1} . M_x . \frac{x}{r^3} = -\frac{\partial}{\partial x}\left(\frac{M_x}{K_1 r}\right),$$

which is the correct potential for an electrostatic doublet of moment M_x in an isotropic medium of dielectric constant K_1. The vector defined by equation (54) is not perpendicular to the radius vector. The H field can be calculated immediately from equation (51), if it is wanted.

If we put $p=0$ in equation (54), we get the electrostatic field due to the doublet at rest.

Field due to M_z. In this case

$$A_x=0,$$
$$A_y=0,$$
$$A_z=\frac{1}{P}\cdot\frac{ipM_z}{\beta^2 c}\cdot e^{ipt}$$

and
$$\Phi=\frac{1}{K_3}\cdot M_z\cdot\frac{z}{P^3}$$
$$=-\frac{\partial}{\partial z}\left(\frac{M_z}{K_3 P}\right) \qquad \ldots\ldots(55),$$

and therefore
$$\left.\begin{aligned}E_x&=-\frac{1}{K_3}\cdot M_z\cdot\frac{\partial}{\partial x}\left(\frac{z}{P^3}\right)e^{ipt}\\ E_y&=-\frac{1}{K_3}\cdot M_z\cdot\frac{\partial}{\partial y}\left(\frac{z}{P^3}\right)e^{ipt}\\ E_z&=-\frac{1}{K_3}\cdot M_z\cdot\frac{\partial}{\partial z}\left(\frac{z}{P^3}\right)e^{ipt}\end{aligned}\right\} \qquad \ldots\ldots(56).$$

If $p=0$ in equation (56) we get the electrostatic field for this dipole, at rest. Since $A_x=0=A_y$, the magnetic lines of force corresponding to this field are circles lying in planes perpendicular to Oz and having their centres on the Oz axis.

It is evident from these solutions that an electric dipole always generates a spheroidal wave. It will not generate the spherical wave if M_x is zero, i.e. if the dipole lies wholly along the Oz axis, so that such a dipole generates a purely spheroidal field. One surmises that a purely spherical field would be generated by a magnetic dipole lying wholly along Oz, and this proves to be the case.

Equation (13c) is consistent with $A_z=0$ if

$$v_z=0$$

and
$$\frac{\partial A_x}{\partial x}+\frac{\partial A_y}{\partial y}=0,$$

i.e.
$$A_x=-\frac{\partial F}{\partial y}$$
$$A_y=+\frac{\partial F}{\partial x},$$

F where
$$F=F(x, y, z, t).$$

Also, by equation (13b), $\quad -\frac{1}{a^2}\ddot{A}_y+\Delta A_y=0, \quad$ since $v_y=0$,

hence
$$-\frac{1}{a^2}\ddot{F}+\Delta F=0,$$

giving
$$A_x=-\frac{\partial}{\partial y}\left\{\frac{1}{r}u\left(t-\frac{r}{a}\right)\right\}$$

and
$$A_y = +\frac{\partial}{\partial x}\left\{\frac{1}{r}\,u\left(t-\frac{r}{a}\right)\right\},$$

i.e.
$$H_x = -\frac{\partial^2}{\partial z\,\partial x}\left\{\frac{1}{r}\,u\left(t-\frac{r}{a}\right)\right\},$$

$$H_y = -\frac{\partial^2}{\partial z\,\partial y}\left\{\frac{1}{r}\,u\left(t-\frac{r}{a}\right)\right\},$$

$$H_z = +\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\left\{\frac{1}{r}\,u\left(t-\frac{r}{a}\right)\right\},$$

which is easily shown to be the field due to a magnetic dipole along Oz at the origin*. A field due to electric and magnetic dipoles along the axis of symmetry Oz consists of physically distinct fields, a radial spheroidal field due to the electric dipoles and a radial spherical field due to the magnetic dipoles. The field due to an isolated oblique dipole cannot be generated by a combination of electric and magnetic dipoles along Oz, for the compensating waves do not appear in either of these special cases.

The spheroidal isophasic is given by

$$t-\sqrt{(\xi^2+\eta^2+z^2)}/a = \text{a constant, say zero,}$$

where
$$\xi=\beta x,\quad \eta=\beta y,\quad \beta^2=K_3/K_1;$$

so that
$$K_3\,(x^2+y^2)+K_1 z^2 = C^2 t^2 \qquad (57),$$

where C is the velocity of light in vacuo, or, alternatively,

$$\frac{x^2}{c^2}+\frac{y^2}{c^2}+\frac{z^2}{a^2}=t^2 \qquad \ldots\ldots(58),$$

i.e.
$$\frac{x^2}{c^2}+\frac{y^2}{c^2}+\frac{z^2}{a^2}=1,\ \text{for } t=1 \qquad \ldots\ldots(59),$$

which is precisely Fresnel's form.

The Fresnel spheroid is, then, a genuine isophasic for a divergent wave, even close to the origin. The isophasics of the spherical field are obviously the spheres

$$r=a \qquad (60),$$

so that the Fresnel surface

$$(r^2-a^2)\,(a^2x^2+a^2y^2+c^2z^2-a^2c^2)=0$$

is a true isophasic for a wave diverging from a point source in a uniaxial crystal. The spheroids (58) are such that, if two are drawn for values of t, say t' and t'', then the lengths of the radii r' and r'' drawn in the same direction are given by

$$\frac{r''}{r'}=\frac{t''}{t'} \qquad \ldots\ldots(61),$$

a relation true for any direction. Consequently the velocity of the spheroid in any fixed radial direction is constant, but of course it is different for different directions. The radii of

$$\frac{x^2}{c^2}+\frac{y^2}{c^2}+\frac{z^2}{a^2}=1=F\,(x, y, z) \qquad \ldots\ldots(62)$$

---

* Cf. the standard Hertz problem in an isotropic medium.

$\mathcal{V}$ consequently give the ray speed $\mathcal{V}$ for point-source spheroidal fields, at points
V sufficiently far from the origin, for we have shown that the Poynting vector is then radial. The sheet itself advances everywhere normally to itself at a speed V which varies from point to point in the sheet. Equation (58) is

$$F(x, y, z, t) = 0,$$

so that the speed is given by

$$|V| = -\frac{\partial F/\partial t}{\{\Sigma\,(\partial F/\partial x)^2\}^{\frac{1}{2}}}$$

$$= \frac{1}{\{(x_0^2 + y_0^2)/c^4 + z_0^2/a^4\}^{\frac{1}{2}}},$$

P when $t = 1$, at the point $P(x_0, y_0, z_0)$.

The tangent plane at P is

$$\frac{x_0 X}{c^2} + \frac{y_0 Y}{c^2} + \frac{z_0 Z}{a^2} - 1 = 0 \qquad \text{......(63)},$$

and the perpendicular distance from O to this plane is

$$\frac{1}{\{(x_0^2 + y_0^2)/c^4 + z_0^2/a^4\}^{\frac{1}{2}}},$$

i.e. the phase speed at P (defined as the speed at which the isophasic advances normally to itself) is equal to the perpendicular distance from O on to the tangent plane at P, i.e. the usual Fresnel rule holds good, and

$$\mathcal{V} \cos \widehat{rn} = V.$$

§ 6. CONCLUSIONS

(i) The Fresnel surface

$$(r^2 - a^2)(a^2x^2 + a^2y^2 + c^2z^2 - a^2c^2) = 0$$

is the true isophasic system ($t = 1$) for divergent light radiating from an electric dipole with its axis in any direction, in a uniaxial crystal.

(ii) Provided we are far enough from the origin for the space derivatives of $e^{-i\alpha r}$ to swamp the space derivatives of such quantities as $\frac{xz}{\rho^2} \cdot \frac{1}{r}$ or $\frac{1}{r}$, then the direction of the Poynting vector for each wave system is radial and the **E** and **H** of each wave system lie in the plane perpendicular to the radius vector and are perpendicular to each other, i.e., $\mathbf{E}_1 \perp \mathbf{H}_1$ and $\mathbf{E}_2 \perp \mathbf{H}_2$; but $\mathbf{E}_1$ is also perpendicular to $\mathbf{E}_2$. For the spherical wave, **H** lies in a meridian; for the spheroidal wave, **E** lies in the meridian plane. For the spheroidal wave, $\mathbf{E} \perp \mathbf{R}$ the radius vector, and therefore does not lie in the tangent plane, in which both **H** and **D** lie. Since the two systems of vectors are cross-perpendicular, i.e. $\mathbf{E}_1 \perp \mathbf{E}_2$ and $\mathbf{H}_1 \perp \mathbf{H}_2$, the resultant Poynting vector is merely the sum of the Poynting vectors for each system.

(iii) The well-known Fresnel construction for finding the direction of **D**, and hence the plane containing **E**, for the spheroidal field, and also the relation between the ray speed $\mathcal{V}$ and the phase speed V apply at points sufficiently far from the origin.

(iv) Fresnel's rules do not apply quite close to the origin, $r \ll \lambda$, for the wave system is not really established in this region. Here the electric field is the electrostatic field due to the dipole, varying sinusoidally with time without appreciable phase lag due to distance.

(v) If the spherical and spheroidal waves are to be considered separately we must exclude points on the z axis, for each of these waves separately gives indeterminate fields on the z axis; for example, H_x by formula (33) contains the factor $\cos\theta . \cos\phi . \sin\phi$, and if we let $\theta \rightarrow 0$ the value of H_z is not unique because it depends on the meridian along which we have approached the point on the z axis. The whole field due to both waves is, however, perfectly determinate at all points except the origin. This is an interesting instance of a case where the actual field cannot be completely resolved, everywhere, into two waves, a spherical one and a spheroidal one. The formulae (53) and (55) show that, when these equations are taken jointly, the whole electric field given by them is precisely what it should be close to the origin.

§7. PRACTICAL DETAILS

In order to produce the rings and brushes as simply as possible the writer has found the following modifications in a Reichart microscope to be sufficient.

(i) A brass ring was turned which fitted into the sub-stage in place of the ordinary condenser fitting. Into one end of this ring the polarizing prism was fitted, and into the other screwed end the usual condenser lens was screwed. This condenser illuminated the bottom surface of a specimen of crystal with a spot of highly convergent polarized light. Such a source is, of course, not the same as an ideal electric dipole source.

(ii) A lens was fitted into an adapter carrying an analysing prism and this adapter was screwed into the end of the draw-tube in place of the usual stop. This arrangement merely converted the draw-tube into a weak microscope for viewing the back of the objective and passing only polarized light; see figure 3. With this arrangement the usual rings and brushes are seen brilliantly both with uniaxial material (quartz and calcite) and biaxial material (arragonite). The most convenient objective was the No. 3 Reichart ($\frac{2}{3}$ in.).

If the draw-tube is withdrawn altogether, the rings and brushes are visible quite clearly to the naked eye if a nicol prism is held in front of the eye and crossed. They appear on the back of the object glass. In the specimens examined, the uniaxial material was cut perpendicular to the optical axis and the biaxial material, perpendicular to the bisector of the axes. The detailed theory of the ring-formation would be difficult for these conditions of illumination, since the isophasic emerging from the crystal into air is not even spherical for the ordinary wave and is still less so for the extraordinary wave. These surfaces, outside the crystal, are not congruent, whereas with plane waves, they are. But symmetry alone is sufficient to secure the formation of the rings. It follows that if the wave-retardation is sufficient to cause blackness with crossed nicols at the point r, ϕ it will do so, whatever ϕ is, in the case of a uniaxial material cut perpendicular to the optical axis, and at $-r, \phi$ in the case of a biaxial material cut perpendicular to the bisector of the optical axes.

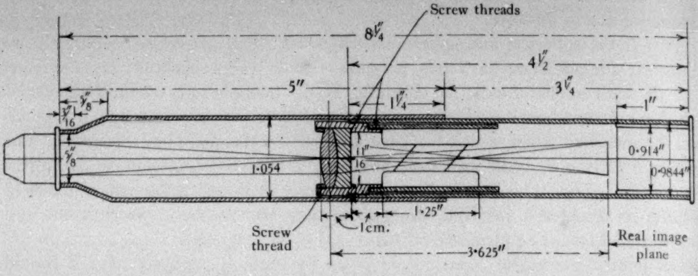

Screw threads
8¼″
4½″
5″
3¼″
⅝″
1¼″
1″
3/16″
⅝″
1·054
11″/16
0·914″
0·9844″
1·25″
1cm.
Screw thread
3·625″
Real image plane

Figure 3.

§ 8. ACKNOWLEDGMENT

In conclusion, it is a pleasant duty to express my gratitude to Dr John Dougall. He has not only undertaken the tedious work of checking the calculations but has made many very helpful criticisms and suggestions.

REFERENCE

(1) Riemann-Weber. *Differentialgleichungen der Physik*, 2, 421, § 4.

621.396.11 : 551.51.053.5

THE PROPAGATION OF MEDIUM RADIO WAVES IN THE IONOSPHERE

BY D. F. MARTYN, PH.D., A.R.C.SC., F.INST.P., Research Physicist,
Australian Radio Research Board

Communicated by Prof. O. U. Vonwiller, July 9, 1934. *Read in title November* 2, 1934.

ABSTRACT. All the available measurements of sky-wave intensities at medium frequencies are collated and expressed as field-strength, distance curves for six typical wavelengths and for distances from 25 to 1000 km. It is shown how this material may be used for the determination of the non-fading radii of broadcasting emitters over country of any effective conductivity. From the observational material an empirical expression for the reflection coefficient of the lower E layer of the ionosphere is derived. It is shown that the observations are incompatible with the existence of a linear or parabolic gradient of ionization in this layer. This incompatibility is not removed by the assumption of an absorbing or D region below the E layer, or by consideration of the variation with height of the collision frequency ν of an electron with the air molecules in the E layer. It is found that the observations can be fully explained if the gradient of ionization is given by the exponential form $N = \epsilon^h$, where h is the height in km. above the region where ionization first becomes appreciable. This gradient also gives rise to equivalent heights which are in agreement with experience. It is found that ν has a value of 10^6 collisions per second at a height of 90 km., in close agreement with Chapman's recent estimate. It is shown that the conclusions reached are not affected by use of the ray methods of geometrical optics, or by neglect of the influence of the earth's magnetic field.

§ 1. INTRODUCTION

A LARGE number of measurements have been made in recent years of the field-intensities due to stations emitting on broadcasting frequencies. In general this field is composed of two parts, that due to the ground-propagated wave and that due to the downcoming wave produced by reflection or refraction in the ionosphere. The propagation of the former wave is reasonably well understood, principally owing to the work of Watson(27), Sommerfeld(26) and T. L. Eckersley(12). The applicability of the analyses of these authors has been confirmed by a large number of measurements accumulated in all parts of the world during the last few years.

Until quite recently few data had been published which would permit a full analysis of sky-wave-propagation at these frequencies. The early work of Appleton and Ratcliffe(5) has now been supplemented by the measurements of the Union Internationale de Radio-Diffusion(11) in Europe, the Federal Radio Commission(23) in America, and the Radio Research Board in Australia(8, 17).

In this paper it is proposed first to review these measurements and to set out curves depicting the variation of field-intensity of the sky wave at distances up to 1000 km. over a frequency spectrum ranging from 1500 to 150 kc./sec. Secondly, this material will be utilized to derive information regarding certain aspects of the structure of the *E* layer of the ionosphere, notably the ionization gradient, the value of the collision frequency of an electron with the air molecules, and the possibility of the existence of an absorbing region at a level below the *E* layer.

§ 2. DISCUSSION OF FIELD-INTENSITY DATA

Examination of the large number of measurements cited above reveals good agreement between the field-intensities of the sky wave obtained in Europe, America and Australia during 1930 and 1931, years of slightly less than average sunspot activity. There is evidence(22) that field-intensities at these frequencies decrease with increasing sunspot activity, so that we may expect sky-wave intensities to reach their maximum values this year. There is not sufficient evidence, however, to indicate whether there will be a differential influence over the radio-frequency spectrum under consideration. On the whole, it seems probable that variation of sunspot activity will only affect the absolute values of the sky-wave intensities, and will not influence appreciably the variation of intensities with frequency and distance from the emitter.

On all frequencies there is evidence of a steady increase of intensity with distance from the emitter, extending up to distances of about 600 km. Thereafter the field decreases steadily, and beyond 2000 km. it appears to fall off according to an inverse-distance law.

For distances less than about 1200 km. there is evidence that the higher frequencies are rather less attenuated than the lower. In this region, however, the ratio of the field-intensities at the extreme ends of the frequency spectrum is seldom greater than 2 : 1, and at greater distances than 1200 km. the attenuation appears to be independent of frequency.

The sky wave fluctuates rapidly from minute to minute and from night to night, so that it is obviously a matter of some difficulty to select from the measurements a quantity which will be truly representative of the sky-wave field-intensity. It has been suggested by a committee at the Lucerne conference that quasi-maximum intensities be employed for this purpose(24).

The quasi-maximum intensities are defined as those values of field-intensity which are exceeded during only 5 per cent of the time of observation. With this definition the sky wave reaches an intensity of 0·35 mV./m. at a distance of 500 km. on a frequency of 1500 kc./sec. for 1 kW. of radiated power, while all frequencies give a value of about 0·04 mV./m. at a distance of 2000 km.

§ 3. AN EMPIRICAL FORMULA FOR SKY-WAVE FIELD-INTENSITIES

As a preliminary to theoretical investigation an attempt has been made to develop a single empirical formula which would express adequately the variations of the observed intensities over the range of observation. If we assume that the radiating properties of the emitting aerial in the vertical plane may be expressed by a cosine law, i.e. that the energy radiated at an angle θ with the horizontal is proportional to $\cos\theta$, then the sky-wave field-intensity E at a distance R km. in the horizontal plane is given by$^{(12)}$

θ

E, R

$$E=\frac{3\sqrt{(10\,W)}\,.\,r\cos\theta}{\sqrt{(R^2+4H_0^2)}}\ \text{mV./m.} \qquad \ldots\ldots(1),$$

where W watts is the total power radiated, H_0 is the height of the reflecting layer in kilometres, and r is the reflection coefficient of the layer.

W, H_0

r

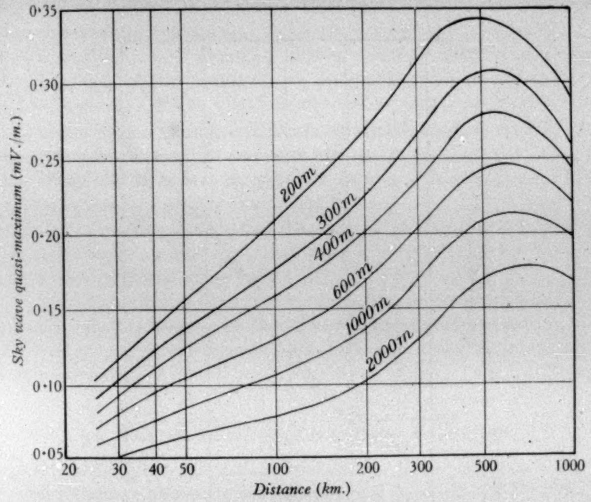


Figure 1.

If the power radiated be 1 kW. then

$$E=\frac{0{\cdot}3rR}{R^2+4H_0^2}\ \text{V./m.} \qquad \ldots\ldots(2).$$

Practically all evening transmission of the sky wave at these frequencies is due to the E layer of the ionosphere, so that without incurring much error we may assign to H_0 an average value of 100 km., and then the only unknown quantity in this equation is r. If now we write for r the empirical expression

$$r=3{\cdot}3\epsilon^{-\cos^2(i_0-35^\circ)\log_{10}\lambda} \qquad \ldots\ldots(3),$$

where

$$i_0=\frac{\P}{2}-\theta,$$

i_0

and λ is the wave-length in metres, then it is found that equation (2) gives a very close approximation to the observed field-intensities. The intensities calculated from equations (2) and (3) are shown in figure 1 for six typical wave-lengths and for

λ

distances ranging from 25 km. to 1000 km. Beyond about 1000 km., radio transmission at these frequencies must occur chiefly by multiple reflections between the earth and the ground, so that the equations cannot be expected to apply without modification.

The curves in figure 1 may be used to determine the distance from a broadcasting station at which fading first becomes serious. This happens when the quasi-maximum sky-wave intensity is equal to one-half the ground-wave intensity. It follows that if ground-wave field-intensity, distance curves be superposed on figure 1, and if an average ground conductivity of 10^{-13} e.m.u. be assumed, with a scale of field-intensities double that shown for the sky wave in figure 1, then the fading-radius is given by the intersection of the ground-wave and sky-wave curves for the desired wave-length λ. For ground of conductivity other than 10^{-13} e.m.u., it is only necessary to substitute for the ground-wave curve of wave-length λ the λ' curve of equivalent$^{(3)}$ wave-length λ', and to utilize the intersection of this curve with the sky-wave curve of wave-length λ. In this way a curve has been derived showing the variation of fading-radius with frequency, for country of average conductivity. It is shown in figure 2 as curve I. In the same figure, for comparison purposes, there are given curves which summarize the experience of broadcasting authorities. Curve II is due to the Union Internationale de Radio-Diffusion, curve III to the Federal Radio Commission, and curve IV to the Van der Pol committee.

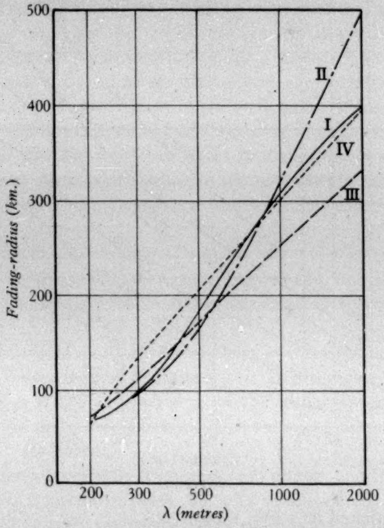


Figure 2.

It will be observed that curve I, which is derived from the above empirical formula for the sky-wave intensities, is in good agreement with the combined experience of these authorities.

Reverting to equation (3), it will be observed that the apparent reflection coefficient of the *E* layer attains a minimum value when the rays are incident at an angle of 35°. It is important to enquire whether this is a true indication of the variation of reflection from the layer with changing angle. Now in any formula of the type (3), if the angle 35° be omitted it is found that (2) gives a maximum of signal-strength at distances considerably less than the observed value of 500 km. The main object of the introduction of this angle is therefore to retard the increase of reflection coefficient with distance. Now in deriving equation (2) we have assumed that the polar radiation diagram of the emitting aerial in the vertical plane is given by a cosine law. There is little doubt that the majority of the aerials used in the actual measurements do not conform to this condition, since most aerials used for broadcasting are designed to radiate as strongly as practicable in the horizontal direction. They will therefore be expected to give a maximum sky wave at a greater distance from the emitter than would an aerial conforming to the cosine law. We therefore attribute the presence of the angle 35° in the formula to this circumstance of aerial design, and consider that it is not necessarily significant in respect of the reflection coefficient of the layer.

In applying these observations to the investigation of the structure of the ionosphere, care has therefore been taken to utilize only measurements made at the longer distances, where uncertainty regarding the high-angle radiation of the aerial becomes unimportant.

§ 4. THEORIES OF SKY-WAVE PROPAGATION

Appleton(2) has made a theoretical investigation of the variation of equivalent height with frequency and angle of incidence, and Appleton and Ratcliffe(5) have examined the dependence of the reflection coefficient of the *E* layer on the same quantities.

They have considered the two cases of a linear and parabolic gradient of ionization in the layer. In both cases it appears that h', the contribution to the equivalent height made by the path in the layer, should vary as some power of $\cos i_0$ and of p, the angular frequency of the wave. The same remarks apply to the theoretically deduced attenuation coefficients of the layer. The same workers have examined these problems experimentally, although over a somewhat limited range of variation of $\cos i_0$ and p. They found, contrary to expectation, that the attenuation coefficient of the layer was nearly constant for different angles of incidence of the wave, while the equivalent height appeared either to be constant or to increase slightly with increasing angles of incidence. In order to explain their results these authors have postulated the existence of a *D* region of ionization below the *E* region. They point out that a ray approaching the layer at a large angle of incidence will have travelled further in the *D* region and experienced greater absorption than one approaching at vertical incidence. In this way the theoretical decrease of attenuation in the E layer with increasing angle of incidence is offset.

Now it is easy to show that the balancing of these opposing tendencies can only be effective at a particular value of i_0, and moreover that this balance will be somewhat critical since it involves the third or fourth power of $\cos i_0$. Moreover for values of i_0 greater than the critical balancing value we should have an increase of attenuation with increasing i_0, a condition which is opposed to the facts revealed in § 3 above.

Again, these authors have suggested that the effect of the D region would be to increase the equivalent height of the E layer for increasing angles of incidence, owing to the "longer path in the D region." But if the equivalent path in the D region is $2D'$ for vertical incidence, then the contribution to the equivalent height of the E region is D', while in the same way, for angle of incidence i_0, the contribution of the D region to the equivalent path is $2D'/\cos i_0$, and to the equivalent height is $2D' \cos i_0/2 \cos i_0$ or D'. It seems therefore that the D region cannot cause a variation of E-layer heights for varying angles of incidence in the manner suggested by Appleton and Ratcliffe. D'

There is, however, another possible way in which the supposed D region could affect the E-layer equivalent heights. In its passage through the D region any ray must be diverted from its original line of travel. Simple optical considerations show that in such an event we have approximately

$$H' = H_0' + D\delta\mu/\cos^2 i_0,$$

H' where H' is the observed equivalent height,

H_0' H_0' the equivalent height which would be observed in the absence of the D region,

D D the thickness of the D region, and

μ $(1-\delta\mu)$ the refractive index of the D region.

It will be observed that the influence of the D region is to cause an increase of H' for increasing values of i_0. Now in Appleton and Ratcliffe's experiments $\cos^2 i_0$ varied from 1 to 0·75 and there was some evidence of an increase of H' with i_0. On the other hand, in some experiments conducted by the Radio Research Board in Australia and shortly to be published, it has been possible to extend the range of variation of $\cos^2 i_0$ from 1 to 0·25, and no marked variation of H' has been found. Now if the D region is responsible for the small variation of H' observed by Appleton and Ratcliffe, then we should expect the variation in the latter experiments to be three times as great, and consequently to be quite marked.

Summing up the above considerations, with regard to both the apparent absorption coefficients and equivalent heights of the E layer, it appears that the hypothesis of a D region is inadequate to explain the results. In what follows an attempt will be made to interpret the results, taking into account possible gradients of ionization in a single E layer and the probable variation of the collision frequency ν with height. ν

§ 5. THE STRUCTURE OF THE *E* LAYER

It appears that the maximum ionization-density $N_{\mathrm{max.}}$ in the *E* layer is attained around noon, typical values being of the order 2×10^5 electrons per cm.^{3}(4). This value decreases steadily until about two hours after sunset, when the rate of decrease falls off considerably, leaving an almost steady ionization of the order of 10^4 electrons per cm.^{3} The latter figure is in good agreement with values obtained in Australia(21). It appears that the height at which this maximum value exists is about 100 km. at noon, and somewhat higher during the night.

Our knowledge of the ionization-gradient is much less definite. For values of N approaching $N_{\mathrm{max.}}$ from below there is some experimental(1) and theoretical(9) evidence to show that the gradient is approximately parabolic. There appears to be no reliable experimental evidence yet available concerning the gradient for small values of N. During night hours the position is even more obscure, there being evidence that the theory of Chapman is then invalid for the lower part of the *E* layer(29).

We shall therefore endeavour to make use of the observations quoted in §§ 2 and 4 above in order to derive information about the ionization-gradient in the lower part of the layer, since it is just this region that is responsible for the propagation of waves of medium radio frequencies. There is considerable evidence that the thickness of the *E* layer is large compared with the wave-lengths we are considering. Now it is not usual for these waves to penetrate the *E* layer, so that in general we shall be justified in neglecting the effect of electron-limitation on the intensity of the waves, and can confine ourselves to the consideration of absorption.

At present our knowledge of the value of ν at the level of the *E* layer is derived from three sources. These are (i) the kinetic theory of gases as applied to the earth's atmosphere, (ii) the observations of Lindeman and Dobson(18) on meteors, and (iii) the theory of Bailey and Martyn(6) on the interaction of radio waves. All three sources give a value for ν of $n\times 10^5$ per second, where n is a small integer. It is one n
of the objects of the present investigation to attempt to evaluate n more accurately. The variation of ν with height is given by the law

$$\nu=\nu_1\epsilon^{-h/H},$$

where ν_1 is the value of ν at the beginning of the *E* layer, ν_1

h the height above this level, and h

H the height of the homogeneous atmosphere. H

The quantity H is a slowly varying function of h but may safely be assumed to be constant over the region we shall consider. Its absolute value depends on the temperature and constitutes of the atmosphere in the *E* layer. It is probable that convection(20) exists at these levels, so that the composition of the atmosphere is not very different from that at the ground. The experiments of Whipple(28) on the refraction of sound-waves in the upper atmosphere, and the calculations of Gowan(14) taking account of absorption in the ozone layer, are in agreement with a temperature of 300° K. Taking these considerations into account, H cannot differ greatly from 10 km. at the level of the E layer.

Evidence has been obtained by Green$^{(15)}$ that two layers of ionization may exist simultaneously at heights near to 100 km. This stratification of the E layer has also been observed recently by Ratcliffe and White$^{(25)}$ on wave-lengths of 75 m. and 150 m. For the wave-lengths and the angles of incidence which we are considering here, there is little doubt that reflection normally occurs from the lower of these two layers. It will be understood therefore that the conclusions reached below apply to this lower layer.

§ 6. STATEMENT OF THE PROBLEM

All the available evidence$^{(5, 21)}$ shows that long-distance transmission up to about 1000 km. is mainly due, at medium frequencies, to a wave which has undergone a single reflection from the E layer. The empirical formula developed in § 3 above therefore gives the actual reflection coefficient of the layer for the ranges of frequencies and angles of incidence specified. Appleton and Ratcliffe$^{(5)}$ have examined this reflection coefficient for the two cases of a linear and a parabolic gradient of ionization in the layer. They obtained a reflection coefficient equal to $\exp\,(-A\nu \cos^3 i_0/\lambda^2)$ in the first case, and to $\exp\,(-B\nu \cos^2 i_0/\lambda)$ in the second, where A and B are constants. In each case a ray theory of refraction was used, and the influence of the earth's magnetic field and the variation of ν with height were neglected. It will be observed that both of these expressions for the reflection coefficient of the layer are in serious disagreement with the experimental results quoted in § 3, which show a small *decrease* in reflecting power with increasing wave-length. This discrepancy has been noted by Appleton and Ratcliffe in examining their own experimental results, and they have attributed it to the presence of an absorbing D region below the E layer, and/or to the fact that ν varies with height.

It has been shown in § 4 that the existence of a D region during the night is incompatible with the evidence. The remaining alternative is examined in the following sections, where account is taken of the variation of ν with height.

In what follows we shall use a ray treatment, and neglect the influence of the earth's magnetic field, two approximations which require justification. It is well known that the methods of geometrical optics are only applicable when the change of optical properties of the medium in the course of a wave-length is small. This condition, which is of fundamental importance in the theory of wave mechanics, has been expressed by de Broglie$^{(10)}$ in the form

$$\mu \cos i \,\frac{d\left(\dfrac{1}{\mu}\right)}{dh} . \lambda \ll 1 \qquad \text{......(4)},$$

h where μ is the refractive index of the medium and h is the direction of maximum gradient in the layer. It will be observed that, for a given wave-length, this condition is most likely to be satisfied for long-distance transmission, where $\cos i$ is necessarily small. We shall then be justified in employing a ray treatment so long as the relation (4) is satisfied.

Turning now to the second proposed approximation we note that for long-distance propagation such as we are considering, in moderately high latitudes, transmission occurs approximately at right angles to the earth's magnetic field. Now in such event the magneto-ionic$^{(3)}$ theory shows that, in the now doubly-refracting layer, the ordinary ray will be propagated as if no magnetic field were present. For the lower wave-lengths the extraordinary ray will then be strongly absorbed and need not be considered further. On the other hand, for the longer wave-lengths in the broadcasting spectrum, the extraordinary ray may have a much smaller absorption coefficient than the ordinary ray, and at first sight it might seem that it could not be neglected. Closer analysis reveals however that the extraordinary wave must penetrate more deeply into the layer, and for likely conditions of ionization-gradient will be absorbed to much the same extent as the ordinary wave. Again, the polarization of the extraordinary wave is such that the magnetic vector is nearly vertical, so that it will produce but a small effect in an aerial at the surface of the ground.

Summing up these considerations we find that we shall be justified in neglecting the influence of the extraordinary ray in long-distance reception, and that we shall obtain a close approximation to the intensity of the ordinary wave by neglecting the earth's field entirely.

§ 7. A THEOREM

We shall be concerned below with the general case of a ray incident on the E layer at any angle of incidence. It is convenient to show that the results for any angle may be deduced from those for vertical incidence by a simple substitution.

Lorentz$^{(19)}$ has shown that the absorption by a dispersive medium of a wave of angular frequency p can be expressed by the coefficient p

$$K = \frac{\nu}{2c\mu} \cdot \frac{4\P N e^2}{m\,(p^2+\nu^2)} \qquad \text{......(5)}, \qquad K$$

where the ratio of the intensity of the emergent wave to that of the incident wave is given by

$$\frac{E}{E_0} = \epsilon^{-\int \kappa\, ds} \qquad \text{......(6)},$$

and μ, the refractive index of the medium, is given by

$$\mu^2 = 1 - \frac{4\P N e^2}{m\,(p^2+\nu^2)} \qquad \text{......(7)},$$

and where c is the velocity of light in free space, and e, m are the charge and mass of an electron in e.s.u. c, e, m

Let us further write $\rho^2 = 4\P e^2/m$ ρ

and $N = \psi\,(h)$. ψ

Then the absorption experienced by a wave at vertical incidence is given by

$$2\int_0^{h(\mu=0)} \kappa ds = \int_0^{h_0} \frac{\nu\,(1-\mu^2)}{c\mu}\,ds$$

$$= \int_0^{\psi^{-1}(p^2/\rho^2)} \frac{\nu_1 \epsilon^{-h/H} \rho^2 \psi\,(h)\,dh}{pc\sqrt{\{p^2-\rho^2\psi\,(h)\}}} \qquad \text{......(8).}$$

provided $\nu^2 \ll p^2$.

In the same way, for an angle of incidence i_0, the absorption is given by

$$2\int^{h\,(\mu=\sin i_0)} \kappa ds = \cos i_0 \int_0^{\psi^{-1}(p^2\cos^2 i_0/\rho^2)} \frac{\nu_1 \epsilon^{-h/H} \rho^2 \psi\,(h)\,dh}{cp\cos i_0 \sqrt{\{p^2\cos^2 i_0 - \rho^2\psi\,(h)\}}} \qquad \text{......(9).}$$

It follows that

$$\left(2\int_0^{h\sin i_0} \kappa ds\right)_{p,\,i_0} = \cos i_0 \left(2\int_0^{h_0} \kappa ds\right)_{p\cos i_0,\,0} \qquad \text{......(10),}$$

so that we need only investigate the absorption for vertical incidence, that at incidence i_0 being immediately deducible therefrom.

By similar reasoning it may be shown that the equivalent path P' in the layer for any angle of incidence may be deduced from that at vertical incidence by the relation

$$(P')_{p,\,i_0} = \frac{1}{\cos i_0}\,(P')_{p\cos i_0,\,0} \qquad \text{......(11).}$$

§ 8. ANALYSIS OF LINEAR IONIZATION-GRADIENT

α Let us write $N = \alpha^2 h$

a and $a = p/\rho\alpha$.

η_0 Then for vertical incidence the attenuation η_0 is given by

$$\eta_0 = 2\int_0^{h_0} \kappa ds = \frac{\nu_1}{ac}\int_0^{a^2} \frac{h\epsilon^{-h/H}\,dh}{\sqrt{(a^2-h)}} \qquad \text{......(12).}$$

Writing $h^{\frac{1}{2}} = a\cos\theta$, we have

$$\eta_0 = \frac{2a^2\nu_1}{c}\,\epsilon^{-a^2/2H}\int_0^{\frac{\pi}{2}} \epsilon^{-a^2\cos 2\theta/2H}\,.\cos^3\theta\,d\theta \qquad \text{......(13).}$$

To effect this integration we make use of Sonine's expansion,

$$\epsilon^{z\cos\theta} = I_0\,(z) + 2\sum_{n=1}^{\infty} I_n\,(z)\cos n\theta,$$

where $I_n\,(z)$ is the modified Bessel coefficient given by

$$I_n\,(z) = \sum_{k=0}^{\infty} \frac{1}{\Pi\,(k)\,\Pi\,(n+k)}\left(\frac{z}{2}\right)^{n+2k},$$

and $\Pi\,(k)$ is Gauss's function having the values unity for $k=0$, and $k!$ for positive integral values of k.

Upon performing this integration, we have to a close approximation

$$\eta_0 = \frac{4a^2\nu_1}{3c}\,\epsilon^{-a^2/2H}\left(1 - \frac{3a^2}{10H} + \frac{19a^4}{280H^2} - \frac{47a^6}{5040H^3} + \frac{a^8}{8870H^4}\right) \quad \ldots\ldots(14),$$

and hence for angle of incidence i_0, by (10),

$$\eta_{i0} = \frac{4a^2\nu_1\cos^3 i_0}{3c}\,\epsilon^{-a^2\cos^2 i_0/2H}\left(1 - \frac{3a^2\cos^2 i_0}{10H} + \frac{19a^4\cos^4 i_0}{280H^2} - \frac{47a^6\cos^6 i_0}{5040H^3} + \frac{a^8\cos^8 i_0}{8870H^4}\right) \quad \ldots\ldots(15).$$

If we neglect the variation of ν with height, then $H = \infty$ and the expression reduces to

$$\eta_{i0} = \frac{4a^2\nu_1\cos^3 i_0}{3c},$$

which is the form obtained by Appleton and Ratcliffe in this simple case, and is in disagreement with the observations. We proceed to examine whether the terms introduced by consideration of the variation of ν with height are adequate to explain the discrepancies.

To investigate the dependence of attenuation on frequency we differentiate equation (15) with respect to a, which is proportional to p. Then if the experimental observations are to be satisfied, we must have

$$\frac{d\eta_{i0}}{da} = 1 - \frac{11a^2\cos^2 i_0}{10H} + \frac{99a^4\cos^4 i_0}{280H^2} - \frac{359a^6\cos^6 i_0}{5040H^3} + \frac{a^8\cos^8 i_0}{200H^4} \leqslant 0 \quad \ldots(16),$$

so that

$$\frac{a^2\cos^2 i_0}{H} \geqslant 1{\cdot}35.$$

But $a^2\cos^2 i_0 = h_{i0}$ is the depth of penetration of the ray into the layer, and it may readily be shown that h', the contribution to the equivalent height of the layer made by the path in the layer, is equal to $2h_{i_0}$ when the gradient of ionization is linear. It follows that $h' \geqslant 27$ km.

This condition must be satisfied over the whole frequency spectrum, so that for wave-lengths near 200 m. we must have $h' \geqslant 2000$ km. So large a value for h' cannot exist in practice, since the total equivalent height of the layer is only about 100 km.

We conclude that the ionization-gradient at the under surface of the E layer is not of linear form.

§ 9. ANALYSIS OF PARABOLIC IONIZATION-GRADIENT

Let us write

$$N = \alpha^2 h^2,$$

then for vertical incidence we have

$$\eta_0 = 2\int_0^{h_0} \kappa\, ds = \frac{\nu_1}{ac}\int_0^a \frac{\epsilon^{-h/H}h^2}{\sqrt{(a^2 - h^2)}}\,.\,dh = \frac{a\nu_1}{c}\int_0^{\pi/2} \cos^2\theta\,\epsilon^{-a\cos\theta/H}\,d\theta \quad \ldots\ldots(17),$$

where $h = a\cos\theta$.

Making further use of Sonine's expansion, we find

$$\eta_0=\frac{av_1}{4c}\left(\P-\frac{8a}{3H}+\frac{3\P a^2}{8H^2}-\frac{16a^3}{45H^3}+\frac{5\P a^4}{192H^4}-\frac{8a^5}{525H^5}\right) \qquad \ldots\ldots(18),$$

so that

$$\eta_{i0}=\frac{a\P\nu_1\cos^2 i_0}{4c}\left(1-\frac{8h_{i0}}{3\P H}+\frac{3h_{i0}^2}{8H^2}-\frac{16h_{i0}^3}{45\P H^3}+\frac{5h_{i0}^4}{192H^4}-\frac{8h_{i0}^5}{525\P H^5}\right) \qquad \ldots\ldots(19),$$

where, as before, $h_{i0}=a\cos i$ and is the depth of penetration of the ray. Differentiating with respect to a, we have, if the attenuation does not decrease with increasing wave-length,

$$1-\frac{16h_{i0}}{3\P H}+\frac{9h_{i0}^2}{8H^2}-\frac{64h_{i0}^3}{45\P H^3}+\frac{25\P h_{i0}^4}{192H^4}-\frac{48h_{i0}^5}{525\P H^5}\leqslant 0 \qquad \ldots\ldots(20),$$

so that $$h_{i0}\geqslant 12 \text{ km.},$$

while $$h_{i0}'=\tfrac{1}{2}\P h_{i0}=19 \text{ km.}$$

By the same reasoning as was employed in § 8 above we find that so large a value for h_{i0}' would give rise to much larger variations of the equivalent height of the E layer than are observed in practice, and we conclude that the gradient of ionization in the layer is not parabolic. It is to be observed however that this gradient gives rise to results which are less divergent from the observed results than those obtained with the linear gradient. Moreover, there would appear to be a greater chance of the observations being satisfied by a layer in which the gradient of ionization varies with the depth of penetration of the incident ray. Accordingly, we proceed in the next section to examine the properties of a layer in which the ionization-gradient is exponential in form.

§ 10. ANALYSIS OF EXPONENTIAL GRADIENT

Let us write $$N=\gamma e^{\beta H}$$

and $$k=\rho^2\gamma/p^2.$$

Then for vertical incidence

$$\eta_0=\frac{k\nu_1}{c}\int_0^{h_0}\frac{\epsilon^{(\beta-1/H)h}\,.\,dh}{\sqrt{(1-ke^{\beta h})}} \qquad \ldots\ldots(21),$$

where $$h_0=\beta^{-1}\log_e k^{-1}.$$

Substituting $$ke^{\beta h}=\cos^2\theta$$

we have $$\eta_0=\frac{2\nu_1 k^{1/\beta H}}{c\beta}\int_0^{\cos^{-1}k^{\frac{1}{2}}}\cos^{(1-2/\beta H)}\theta\, d\theta=\frac{2\nu_1 k^{1/\beta H}}{c\beta} \qquad \ldots\ldots(22),$$

provided that $$\left|\frac{2}{\beta H}\right|\ll 1$$

and $$k\ll 1.$$

Hence for angle of incidence i_0 we have

$$\eta_{i0}=\frac{2\nu_1}{c\beta}k^{1/\beta H}\cos^{(1-2/\beta H)} i_0=\frac{2\nu_1\cos i_0}{c\beta}\left(\frac{\rho\gamma^{\frac{1}{2}}}{p\cos i_0}\right)^{2/\beta H} \qquad \ldots\ldots(23).$$

It is seen that the attenuation decreases slowly with increasing frequency, and varies approximately as cos i_0. The dependence of attenuation on these factors is therefore just of the type indicated by experiment. We proceed to examine the magnitude of the quantities involved.

It is clearly necessary first of all to determine the value of γ, which is the ionization-density postulated at the foot of the layer. Now the theory of Lorentz, on which we have based our analysis, becomes invalid when the number of electrons is small in a cube whose side is equal to 1 vacuum wave-length, so that a lower limit exists to the permissible value of γ. At the same time an upper limit is set by the fact that if γ is too great, appreciable absorption will occur below the region defined as the foot of the layer. For the frequencies under consideration we can safely assume a value of unity for γ, so that we consider the layer to commence at the point where one electron per cm^3 is found.

The gradient of ionization is determined by the value of β. Let us take β equal to 1 and examine the magnitudes of the quantities involved. Consider the value 10^7, for which $\lambda = 190$ m. of the frequency p, and the angle of incidence 70°, for which $R = 550$ km.

Then by equation (23),

$$\eta_{70} = 0{\cdot}97 \times 10^{-6}\, \nu_1.$$

But by measurement we have from equation (3), for the average[*] value of η_{70},

$$\eta_{70} = \cos^2 35^\circ \log_{10} 190 - \log_e 3{\cdot}3/2 = 1{\cdot}05,$$

hence the theoretical value of the attenuation is in agreement with the measured value if $\nu_1 = 1{\cdot}1 \times 10^6$. Taking this value for ν_1 we proceed to examine the variation of the reflection coefficient with p. A comparison between the observed and calculated values of η over the complete range of broadcasting frequencies is shown in figure 3. It is seen that the calculated values are in close agreement with the observed values over the entire range.

It will be observed that this comparison has been made for a large value of i_0, where uncertainty regarding the radiation characteristics of the aerials employed is of little importance. The variation of η with i_0 is set out in figure 4, where a comparison is made between the observed and calculated values of the attenuation, for values of R ranging from 350 to 1150 km. The comparison is made for a wave-length of 190 m., and once again it is seen that close agreement exists between the observed and calculated values.

It appears therefore that the facts of long-distance transmission on broadcasting frequencies can be accounted for, to a very close approximation, by refraction in a layer in which the ionization-gradient is given by $N = e^{h(km)}$. The agreement with experience is so good that we proceed to examine this model of the E layer more closely. In the first place it appears desirable to test the applicability of the ray theory to such a layer. Writing

---

[*] The average value of the reflection coefficient of the layer is found to be approximately one-half of the quasi-maximum value. The larger value of the latter coefficient is probably due mainly to the presence of multiply reflected rays from the E layer. The effect of these rays will be largely eliminated by taking the average value of the reflection coefficient.

$$\mu^2 = 1 - \delta\mu$$
$$= 1 - k\epsilon^h,$$

we find that (4) becomes

$$\tfrac{1}{2}\mu \cos i \, d\mu . \lambda \ll 1,$$

where λ is measured in km.

Now $\delta\mu$ has a maximum value of $\cos^2 i_0$ and $\cos i$ varies from $\cos i_0$ to zero along the trajectory of the ray, so that we may safely say that the ray theory is applicable for the ranges of wave-length and distance which we have considered above.

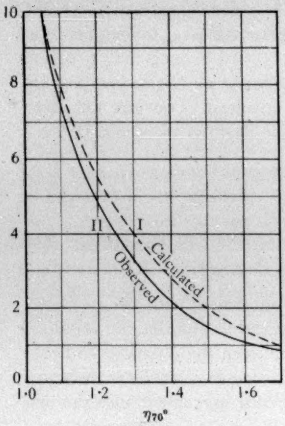


Figure 3.

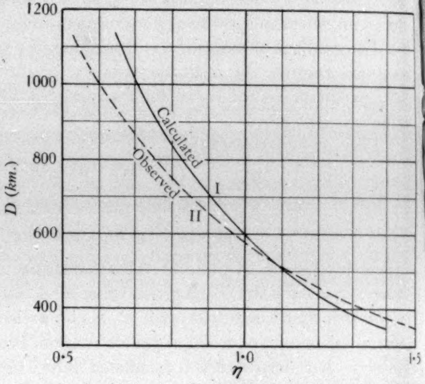


Figure 4.

Secondly, it will be remembered that in § 7 we have assumed that $\nu^2 \ll p^2$. Now we have seen above that ν_1 has the value $1 \cdot 1 \times 10^6$ collisions per second, so it would seem that the above assumption is invalid. It must be pointed out, however, that little absorption occurs in the regions where ν is of the order 10^6, since the ionization there is small. The greater part of the absorption occurs near the top of the ray's trajectory, where ν may be but $\frac{1}{3}$ of this value, so that the condition $\nu^2 \ll p^2$ may be satisfied even for the longest wave-lengths. It is to be noted, however, that even were this assumption to prove invalid for the longest wave-lengths it would lead to but little alteration in our conclusions, while for the shorter wave-lengths no doubt as to its validity exists.

Finally, it is necessary to examine the equivalent heights which would be measured for this layer. For vertical incidence we have

$$h_0' = \int_0^{h_0} \frac{dh}{\sqrt{(1 - k\epsilon^{\beta h})}},$$

and writing $k\epsilon^{\beta h} = \cos^2\theta$, we have

$$h_0' = \frac{2}{\beta}\int_0^{\cos^{-1}k^{\frac{1}{2}}} \frac{d\theta}{\cos\theta} = \frac{1}{\beta}\log_e\frac{4}{k},$$

where $k^{\frac{1}{2}} \leqslant 1$.

Hence for incidence i_0,

$$h'_{i0} = \frac{1}{\beta}\log_e\frac{4\cos^2 i_0}{k},$$

so that, if $$\beta = \gamma = 1,$$

then $$h'_{i0} = \log_e(1{\cdot}3\times10^{-9}p^2\cos^2 i_0);$$

when $p = 10^7$ and incidence is vertical, $h_0' = 12$ km., while when $p = 10^6$ and incidence is vertical $h_0' = 7$ km. These values appear entirely reasonable, although no simultaneous height-measurements with which they could be compared have been made over this range of wave-lengths.

Since the total equivalent height of the layer measured on these wave-lengths is normally near to 100 km. it follows that the value of $\nu_1 = 10^6$ occurs at a height of about 90 km. This result is in very close agreement with Chapman's estimate, as quoted by T. L. Eckersley$^{(13)}$.

The variation of equivalent height with angle of incidence is very small, in agreement with the observations$^{(5, 21)}$.

§ 11. DISCUSSION AND CONCLUSIONS

It has been shown that on medium radio frequencies the intensities of the sky waves, and their variation with frequency and distance, are inconsistent with the presence of a linear or parabolic gradient of ionization in the *E* layer. The assumption of an absorbing *D* region of ionization does not remove the inconsistencies.

On the other hand, the facts of long-distance transmission, where the angle of incidence of the sky wave on the *E* layer is greater than 60°, can be accounted for to a close approximation if the gradient of ionization be of exponential form. It has been explained that the observations at short distances become unreliable owing to the uncertainty regarding the vertical polar diagram of the emitting aerials. Nevertheless it may be shown from equations (2) and (3) that the field-intensities observed very close to the emitter are only a few times greater than those deduced from equation (23). This result seems entirely reasonable in view of the important influence of even a small horizontal portion of the emitting antennae when θ approaches 90°, and the consequent serious departure from the assumed polar radiation diagram which must occur at such short distances. Our inability to use the short-distance observations does not however limit seriously the amount of material with which the theory can be compared. Between 60° and 80°, cos (i_0) varies from 0·5 to 0·17, while the distances covered, from 350 km. to 1100 km., ensure that the results of the great majority of the observing stations are utilized. The range of variation of p is some tenfold.

Again, it should be pointed out that in utilizing the measurements for the deduction of the reflection coefficient of the ionized layer no account has been taken of the imperfect conductivity of the ground and its influence on the measured fields. It may be shown(7) that this effect becomes of importance for the shorter wave-lengths and the longer distances, when it may reduce the observed field-intensities by some 20 per cent. It follows that the reflection coefficient of the layer for the shorter wave-lengths is slightly greater than is indicated by equation (3). It appears therefore that a value of ν slightly less than $1{\cdot}1 \times 10^6$ collisions per second might give an even better fit to the observations. In view, however, of the statistical nature of the observations, and of the necessary variabilities of emitting aerials and observing personnel, as well as the naturally occurring variations in the ionized layer, it does not seem profitable at this stage to attempt a more accurate determination of the values of ν or of the ionization-gradient.

Summing up the discussion and analysis of the wide range of observations available, it appears difficult to escape the conclusion that the gradient of ionization in the layer is very sharp and closely approaches exponential form, that the collision frequency ν has the value 10^6 at a height of about 90 km., and that the *D* region, if it exists, must have a very low ionization content.

As a corollary to these conclusions it follows that the equivalent height of the *E* layer measured at long wave-lengths should not differ by more than a few km. from those measured at the lower end of the broadcasting band of wave-lengths. Experimental evidence on this point appears to be lacking. The measurements of Hollingworth(16) on a wave-length of 14,350 m. show heights of 90 km. during the day in winter, although there is necessarily some doubt as to the accuracy of height-measurements on these very long wave-lengths.

It is, however, a necessary consequence of the conclusions reached in this paper that the equivalent height of the layer for the longer wave-lengths in the medium-frequency spectrum (2000 m.) should not differ greatly from 95 km.

§ 12. ACKNOWLEDGMENTS

The work described has been carried out as part of the programme of the Australian Radio Research Board. The author wishes to express his indebtedness to Mr W. J. Wark, M.Sc., for considerable assistance in the compilation of the field-intensity data; to Dr A. L. Green, who has carefully checked all the calculations and made some valuable suggestions; to Prof. V. A. Bailey for the benefit of some helpful discussion; and finally to the Chairman of the Board, Prof. J. P. V. Madsen, for his continued encouragement and advice throughout the progress of the investigation.

REFERENCES

(1) APPLETON. *Proc. phys. Soc.* **42**, 321 (1930).
(2) APPLETON. *Proc. phys. Soc.* **41**, 43 (1928); **42**, 321 (1930).
(3) APPLETON. *J. Instn elect. Engrs,* **71**, 642 (1932).
(4) APPLETON and NAISMITH. *Proc. roy. Soc.* A, **137**, 36 (1932).
(5) APPLETON and RATCLIFFE. *Proc. roy. Soc.* A, **128**, 133 (1930).
(6) BAILEY and MARTYN. *Phil. Mag.* **18**, 369 (1934).
(7) BOUTHILLON. *J. Éc. Polyt.*, Paris, **25**, 151 (1925).
(8) *Bull. Counc. sci. industr. Res. Austr.* No. 63 (1932).
(9) CHAPMAN. *Proc. phys. Soc.* **43**, 26 (1931).
(10) DE BROGLIE. *J. Phys. Radium,* **7**, 321 (1926).
(11) Documents du C.C.I.R. Deuxième Réunion, Mai–Juin, 1931.
(12) ECKERSLEY. *Proc. Inst. Radio Engrs,* N.Y., **20**, 1555 (1932).
(13) ECKERSLEY. *Proc. roy. Soc.* A, **137**, 169 (1932).
(14) GOWAN. *Proc. roy. Soc.* A, **128**, 531 (1930).
(15) GREEN. *Bull. Counc. sci. industr. Res. Austr.* No. 59 (1932).
(16) HOLLINGWORTH. *J. Instn elect. Engrs,* **64**, 579 (1926).
(17) *J. Instn Engrs Aust.* **5**, 193 (1933).
(18) LINDEMAN and DOBSON. *Proc. roy. Soc.* A, **103**, 339 (1923).
(19) LORENTZ. *Theory of Electrons,* p. 132.
(20) MARIS. *Terrestrial Magnetic and Atmospheric Electricity,* p. 233 (1928).
(21) MARTYN, CHERRY and GREEN. *Proc. phys. Soc.* **47**, 340 (1935).
(22) PICKARD. *Proc. Inst. Radio Engrs,* N.Y., **15**, 749 (1927).
(23) *Proc. Inst. Radio Engrs,* N.Y., **20**, 611 (1932).
(24) Rapport du DR VAN DER POL. *Conférence Européenne des Radiocommunications,* Document No. 1, p. 41 (Lucerne, 1933).
(25) RATCLIFFE and WHITE. *Proc. phys. Soc.* **46**, 107 (1934).
(26) SOMMERFELD. *Ann. Phys.*, Lpz., **81**, 1135 (1926).
(27) WATSON. *Proc. roy. Soc.* A, **95**, 83 (1919).
(28) WHIPPLE. *Quart. J. R. met. Soc.* p. 331 (1931).
(29) WHITE. *Proc. phys. Soc.* **46**, 91 (1934).

621.396.11 : 551.51.053.5

LONG-DISTANCE OBSERVATIONS OF RADIO WAVES OF MEDIUM FREQUENCIES

By D. F. MARTYN, Ph.D., A.R.C.Sc., R. O. CHERRY, M.Sc.
AND
A. L. GREEN, Ph.D., Research Physicists,
Australian Radio Research Board

Communicated by Prof. O. U. Vonwiller, July 9, 1934. Read in title November 2, 1934.

ABSTRACT. The frequency-change technique of Appleton and Barnett has been applied to the analysis of the downcoming waves from a distant transmitter. The observations were carried out simultaneously at distances of 25 and 700 km. from the emitter, which operated on a frequency of 1415 kc./sec. It was found that several downcoming waves were present at the more distant receiving station. Each of these waves was identified by using the path-length of the singly reflected wave from the *E* layer as a reference. In this way it was found that the equivalent heights of both the *E* and the *F* layers are relatively stable over the 700-km. transmission path, and do not vary appreciably with the angle of incidence of the wave. The equivalent height of the *F* layer showed a pronounced minimum at about 3 a.m. each morning. The rate of propagation of the minimum height in the horizontal direction appears to be slower than the rate of sunset propagation in the same direction. The ionization-density in the *E* layer in the early morning was always greater than $2\cdot4 \times 10^3$ electrons per cm.^{3}, and during half the period of the observations was less than $8\cdot3 \times 10^3$ electrons per cm.^{3} The intermediate layer was observed regularly at sunrise. From the measurements of equivalent heights at different angles of incidence it is concluded that the gradient of ionization at the lower boundary of the *E* layer is sharp.

§ 1. INTRODUCTION

In general the field-intensity due to an emitting station is made up of two parts, that due to the ground wave which has been propagated over the earth's surface and that due to downcoming waves which have experienced one or more reflections from the ionosphere. Numerous investigations⁽²⁾ have been made of the characteristics of the downcoming waves at comparatively short distances, of the order 100 km., from the emitter. The particular advantage of this procedure is that the ground wave is then comparable in intensity with the downcoming wave, so that it may be used as a reference wave, of known and stable characteristics, with which the much more variable downcoming wave may be compared. In this way much valuable information has been obtained about the equivalent or virtual heights of the layers of the ionosphere and of the ionization-densities in these layers.

On the other hand it is not possible for such conditions to examine effectively the variation, with changing angles of incidence, of the equivalent heights of the layers and of their reflection coefficients. A knowledge of these variations would be of considerable value in the elucidation of the structure of the ionosphere, and in particular of the gradient of ionization therein. Now, in general, if i be the angle of incidence of the wave on the layer, then the theoretical dependence of the equivalent height and reflection coefficient on i is given by some low power of $\cos i$. For the experimental conditions outlined above the maximum possible change in $\cos i$ is about 15 per cent, so that great refinement of technique would be required to examine the small variations involved.

In the experiments described in this paper two receiving centres were used, at distances of 700 and 25 km. from the emitter, so that the range of variation of $\cos i$ was from 0·3 to 1 or some 200 per cent.

Preliminary observations by one of us(5) indicated that over the greater distance mentioned more than one downcoming wave was receivable. The present series of experiments has established the fact that one of these waves, that which has undergone a single reflection at the surface of the E layer, is invariably present. Moreover, simple geometrical considerations show that the length of the equivalent path of this wave is almost independent of the precise height of the E layer, normally occurring changes in which produce a change of only 1 per cent in the total path. It is possible then to make use of this wave as a reference wave of known path-length and to measure therefrom the equivalent paths of other waves which may be present.

Again, it might be anticipated that a complication would exist in such long-distance experiments in that the constitution of the ionosphere might not be uniform over the entire path. Evidence to the effect that at times this non-uniformity exists has been found in the experiments, but it has been possible to interpret the measurements even in such conditions, and indeed it has been possible to utilize them in order to study the height of the layer at several points above the great-circle path between emitter and receiver, and so to study the horizontal propagation of changes in the layer.

Finally, it has been possible to obtain measurements of the maximum ionization-density in the E layer by noting the presence or absence of the multiply reflected waves from that region. For the frequency employed (1415 kc./sec.) this layer is usually penetrable at vertical incidence during the night, but is seldom penetrated for the larger angles of incidence. In this way it has been possible to study the night-to-night variation of the maximum ionization-density in the layer during the period of the observations.

§ 2. EXPERIMENTAL PROCEDURE

The emitter was located in the P. N. Russell School of Engineering in the University of Sydney. It radiated approximately 1 kW. on a frequency of 1415 kc./sec. The receiving sites were located at Liverpool, N.S.W., and Melbourne, Victoria, distant 25 and 700 km. respectively from the emitter. Throughout the investigation the frequency-change method of equivalent-path determination due to Appleton and Barnett(3) was employed. The frequency-change was normally about 7·5 kc./sec.

For recording purposes an Einthoven galvanometer and camera were used at Liverpool. In Melbourne a Moll galvanometer having a period of 0·2 sec. was used in conjunction with a photographic drum recorder of a type previously employed(5). The frequency-changes, which are automatically produced at the transmitter, occupied about three seconds each. Since the number of interference fringes obtained in Melbourne ranged from 2 to 16 there appeared to be a possibility that the fringes might be distorted owing to the relatively great period of the galvanometer. A calibration of the galvanometer response at low frequencies was made, and a correction factor was found which would enable the true amplitude of any given set of fringes to be deduced. As had been anticipated, this correction factor is larger for the larger fringe numbers. The present paper is concerned only with equivalent-path determinations, which depend only on the number of fringes present, and this correction factor has not been used in it. Indeed, for this work certain advantages accrue from the use of a galvanometer having a comparatively low period, since the fringes due to the higher orders of reflection are relatively suppressed, the interpretation of the records for the lower orders of reflection being thus facilitated. Moreover there is a tendency for the reception of low-amplitude atmospherics to be suppressed, and this again facilitates the interpretation of the records. The galvanometer was shunted with a resistance of 29 Ω., a value which was found to give a satisfactory balance between the undesirable features of over- and under-damping.

The receiver used in Melbourne had three stages of high-frequency amplification employing variable-mu valves. The necessary flatness of the frequency, amplitude response curve was obtained by slightly detuning each stage of amplification. With the screen grid of the detector valve connected to the negative terminal of the high-tension supply the plate current was 2 or 3 μA. when no signal was being received. The rectified current varied linearly with the intensity of the incoming signal, reaching a value of 75 μA. for the maximum signals received, which had a field intensity of the order 0·5 mV./m.

At Liverpool, owing to the strong ground-wave signal, it was found better to use a square-law detector in which the steady plate current had been balanced out. The two stages of radio-frequency amplification were coupled by a band pass unit to ensure flatness of tuning.

The experiments were carried out from midnight to sunrise on eleven mornings between the 5th and 28th October, 1932. At intervals of ten minutes a set of six complete frequency-changes lasting one minute was sent out from the emitter. During the pre-sunrise period, when conditions were changing rapidly, the interval between each set of observations was shortened to five minutes.

§ 3. EXPERIMENTAL RESULTS

General discussion. The results obtained at Liverpool were typical of those obtained for such frequencies over short distances, and call for no special comment here. It need only be remarked that the most frequently observed wave was one which had been reflected once from the *F* layer. On two days out of the eleven the

E layer was observed during the greater part of the morning. On most mornings the intermediate layer was observed during the transfer from the *F* to the *E* region which is associated with the sunrise period.

The results obtained in Melbourne were considerably more complex. On all occasions interference fringes were obtained, showing that at least two downcoming waves were invariably present. A typical set of fringes is reproduced in figure 1, where it is seen that the primary number of fringes is 4·2 while there is evidence of a secondary set of 12·5 fringes. During the series of experiments almost every possible number of fringes between 2 and 18 has been counted, and at first sight the analysis would appear to be difficult. If the fringe-counts be plotted against time, however, it is soon seen that each count lies on one of several continuous curves, and it only becomes necessary to identify the rays corresponding to each curve. The identification was facilitated by the curves set out in figure 2, which show the path-lengths and angles of incidence of the singly, doubly and triply reflected rays plotted as a function of layer-height. We shall denote the singly, doubly, triply, etc. reflected waves from the *E* region by E_1, E_2, E_3, etc. and from the *F* and intermediate regions by F_1, I_1, etc. Then the curve showing the path difference between E_1 and F_1 say, will be denoted by (F_1-E_1).

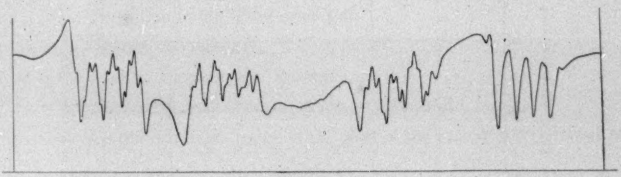


Figure 1.

On plotting the curves it is found that the most commonly observable fringe-counts are produced by (F_1-E_1), (F_2-E_1), (E_2-E_1), (F_2-E_2) in that order of frequency of occurrence. It is seen, therefore, that the principal part of the received signal is due to E_1, the singly-reflected wave from the *E* region. The rays E_2 and F_1 have never been observed simultaneously, but it is frequently noticed that one ray may give place to the other for a brief time. Now reference to figure 2 shows that for normally occurring *F*-layer heights these two rays approach the *E* layer at almost the same angle of incidence. It is clear, therefore, that the transit from the E_2 to the F_1 ray is due to electron-limitation in the *E* layer. The size of the fringes produced by the E_2 and the F_1 rays are usually comparable.

Temporal path variations. When the *F*-layer heights are steady the values measured at Liverpool and Melbourne are remarkably equal. In such circumstances the heights measured in both places do not differ by more than 10 km., which is about the limit of accuracy of the measurements for this layer. The *E*-layer heights vary by not more than two or three km. in similar circumstances.

One of the most striking features of the results is the pronounced minimum which occurs in the *F*-layer equivalent heights at approximately 3 a.m. (Eastern

Australian standard time). This effect, which has been noticed by other workers[8], was found on seven out of nine mornings when the F layer was observable. The layer usually commences to fall at about 1.30 a.m. and after passing through a minimum value attains its normal value again at about 4 a.m. The total fall is usually about 40 km. It is found that the equivalent height of the F layer measured at Liverpool attains this minimum value some time before that measured in Melbourne. If the height in the latter case be deduced from the (F_1-E_1) curve then the difference in

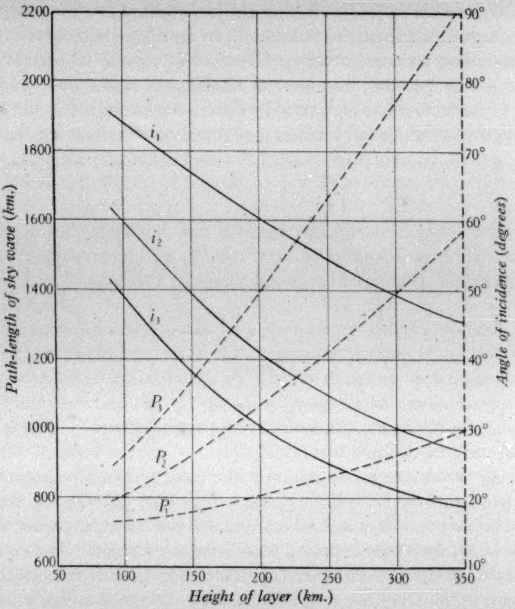


Figure 2.

Δt time Δt between the attainment of the minimum heights is about 30 minutes. Now the difference between the times of sunset in Sydney and Melbourne is 24 minutes, so that if the minimum in the F-layer heights is directly associated with solar influences, then we should expect Δt to have a value of about 12 minutes. It appears, therefore, that the rate of propagation of the F-layer minimum height in the direction Sydney–Melbourne (49° 11′ W. of S.) is slower than the rate of sunset propagation in the same direction.

Shortly before sunrise the F-layer heights recorded at both Liverpool and Melbourne rise to abnormally great values, and eventually E-layer reflection sets in.

There is no doubt that this phenomenon is due to the reduced group-velocity caused by increasing ionization below the F layer.

It is noticeable that the apparent rise in the F layer observed at Melbourne takes place much more slowly than that recorded at Liverpool. This is attributed to the fact that the sun's rays must take considerably longer to irradiate the path of the F_1 ray received in Melbourne than that of the F_1 ray received at almost vertical incidence in Liverpool.

Ionization-densities of the E layer. For the case of propagation between Sydney and Melbourne the angle between the direction of the earth's magnetic field and the direction of propagation of the rays at the apices of their paths in the E layer is such that propagation approximates to the transverse type of the magneto-ionic theory(4). From the theory it may be shown that in this case the ordinary ray* will be much less attenuated than the extraordinary ray, and the refractive index of the layer for the ordinary ray will be given very nearly by

$$\mu^2 = 1 - 4\pi N e^2 / m p^2,$$

where μ is the refractive index of the medium,

N is the density of ionization,

p is the popular frequency of the wave, and

e, m are the charge and mass of an electron (e.s.u.).

By Snell's law the refractive index at the apex of the path of the refracted ray is given by

$$\mu = \sin i,$$

where i is the angle of incidence of the ray on the medium. Hence penetration of the medium occurs when the maximum ionization-density in the medium is less than N_0, where

$$N_0 = \frac{m p^2 \cos^2 i}{4\pi e^2}.$$

The critical ionization-densities for which penetration of the E layer occurs are set out in table 1 for the various rays observed.

Table 1

Ray and wave-length (km.)	i (degrees)	N_0 ($10^3 \times$ electrons per cm.^{3})
E_1	72	2·4
E_2	57	7·4
F_1 (250)	55	8·3
F_1 (300)	49	11
F_2 (250)	35	17
F_2 (300)	30	19
I_2	48	11

Throughout the course of the experiments the ray E_1 was always present. It follows that the ionization-density in the E layer was always greater than $2{\cdot}4 \times 10^3$ electrons per cm.^{3} The F_1 ray was present during 55 per cent of the time of observation, so that the ionization-density in the E layer was less than $8{\cdot}3 \times 10^3$ electrons per cm.^{3} during that time. On two nights the ionization-density was seldom less than 10^4 electrons per cm.^{3}, as evidenced by the almost complete absence of F rays. It is to be remarked that if the Hartree polarization term be proved valid then the densities given above must be increased by 50 per cent.

---

* We have here neglected the controversial Hartree polarization term (*Nature*, **132**, 929 (1933)).

§ 4. ANALYSIS OF TYPICAL RESULTS

Test of October 6, 1932. *Sunrise at Sydney*, 5.27 *a.m.* The results obtained on this morning are shown in figure 3 in which the uppermost curve gives the layer-heights measured at Liverpool, while the lower curves give the path-differences measured in Melbourne. The left-hand scale of ordinates refers to the Liverpool measurements and the right-hand one to those made in Melbourne. It will be seen that the $(F_2 - E_1)$ curve is practically unbroken, indicating the almost continuous presence of the F_2 ray. The minimum path-difference occurs about 27 minutes after the minimum height recorded at Liverpool.

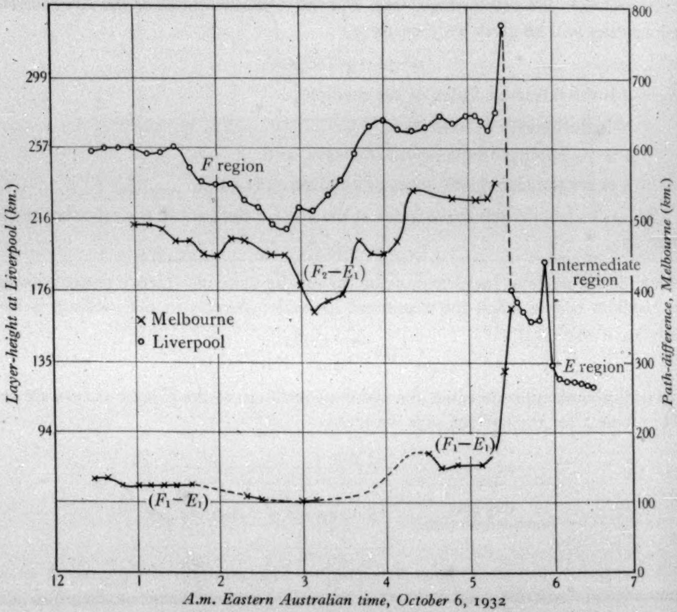


Figure 3.

The (F_1-E_1) curve is more irregular, breaks occurring at 1.40, 2.30, and 3.10 a.m. We interpret these breaks as due to nocturnal increases of the ionization in the E region.

The close agreement of the F-layer heights measured at Liverpool and Melbourne is readily apparent at 1 a.m. and 5 a.m., times when the F layer is relatively stationary. Making use of figure 2, we find the F-layer height* at 1 a.m. to be (*a*) 256 km., (*b*) 252 km., and (*c*) 245 km. as deduced from (*a*) Liverpool observations, (*b*) the difference (F_2-E_1), and (*c*) the difference (F_1-E_1). In the same way we find at 5 a.m. heights of (*a*) 275 km., (*b*) 262 km., and (*c*) 268 km. deduced in the same way. At 5.25 and 5.30 a.m. there is evidence of considerable group-retardation in the F_1 ray, showing that the ionization below the F layer has increased considerably.

Test of October 21, 1932. *Sunrise at Sydney*, 5.07 *a.m.* The results obtained on this morning are shown in figure 4. This test gave perhaps the most interesting results of the present series. During the early hours of the morning the F_1 ray was not received, showing that the ionization in the E layer was greater at that time than in the test described in (*a*) above.

At 1.30 a.m. the E layer is detectable at Liverpool, while from 1.20 a.m. onwards the F_2 ray is received in Melbourne. Now the results obtained in Melbourne fall into two continuous curves, and consideration of figure 2 shows that it is very probable that these two curves are due to (*a*) the difference (F_2- a singly reflected ray from the E region) and (*b*) the difference (F_2- a doubly reflected ray from the E region). Now, if such be the case, then at any time the difference in the ordinates of these two curves should give the value of (E_2-E_1). Between 1 and 2 a.m. the average value of this quantity is 73 km., so that the height of the E region must be near to 100 km. At 2 a.m. however the upper curve starts to rise, while the lower curve starts to fall, and at 2.20 a.m. the difference in the ordinates is 160 km., corresponding to an E-layer height of 150 km., which is a height considerably greater than normal for this layer. There is however another possibility. It has been suggested by Ratcliffe and White(7) that the E layer has a stratified formation, one layer, which they term the e layer, existing at a height of about 105 km., while the other has a height of about 135 km. Now in our experiments an e_2 ray must penetrate the e layer before an e_1 ray, owing to the smaller angle of incidence of the former ray. We consider therefore that up till 2 a.m. both the singly and doubly reflected rays from the lower region come from the e layer, while at 2 a.m. the e_2 ray penetrates to the E layer at about 130 km. At 2.30 a.m. the difference between the ordinates has fallen to a steady value of 100 km., showing that the e_1 ray has now penetrated to the E layer proper.

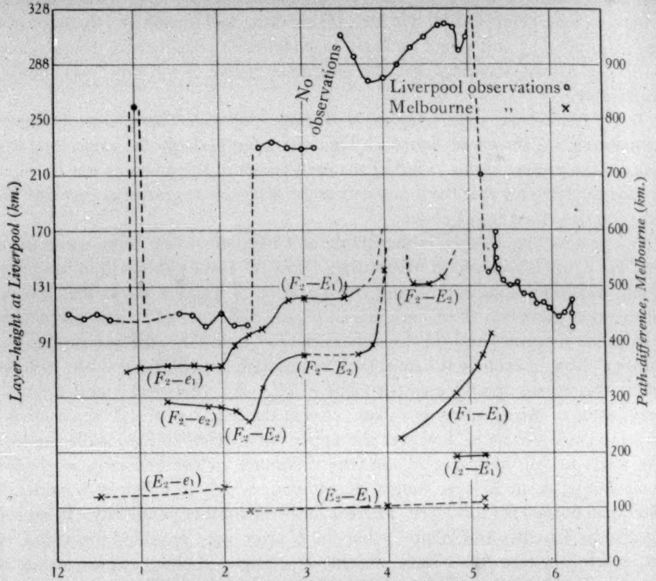
This interpretation of the results is supported by the observation of fringes corresponding directly to (E_2-e_1) at the end of the 2 a.m. record, and again by the observation of a fringe count giving (E_2-E_1) at the end of the 2.20 a.m. record.

Again, Ratcliffe and White have given reasons for believing that the presence of the *e* region is associated with magnetic storms. Reference to the *Cosmic Data Ursigrams* for October 21 shows that this day was marked by moderate disturbance consisting of irregular oscillations. It is worthy of remark that all other days on which tests were conducted were magnetically quiet, and in none of them was there evidence of spasmodic appearance of the *e* layer.

---

* All height-measurements have been corrected in order to allow for the effect of the curvature of the earth.

The *F* layer measured at Liverpool rose sharply in height between 3.10 and 3.30 a.m.* This increase is reflected some 30 minutes later in the Melbourne curves.



A.m. Eastern Australian standard time, October 21, 1932

Figure 4.

After 4 a.m. the fringes become so numerous, and consequently so small, that they could not be counted. Between 3.20 and 3.50 a.m. the *F* layer fell at Liverpool, and this is shown at 4.20 a.m. in Melbourne by the reappearance of the (F_2-E_2) curve.

At 4.10 the F_1 ray made its first appearance in Melbourne and the (F_1-E_1) curve rises steadily until 5.15 a.m. when the *E* layer becomes impenetrable. At 4.50 and 5.10 a.m. there is evidence of a doubly reflected ray from the intermediate region.

Test of October 28, 1932. *Sunrise at Sydney*, 5.00 *a.m.* On this morning, figure 5, the ionization-density was so great that the *E* layer was only penetrable for a few minutes, even for the vertically incident rays at Liverpool. It will be seen that the *e* layer was responsible for the reflection of the rays received in Melbourne till 3.30 a.m., when the e_2 ray penetrated to the *E* region. Almost simultaneously penetration of the *e* layer occurred at Liverpool. At 4.30 a.m. the F_1 ray appeared in Melbourne, and group-retardation set in, causing a rapid increase in the path difference (F_1-e_1).

---

* Unfortunately no observations were made during this interval.

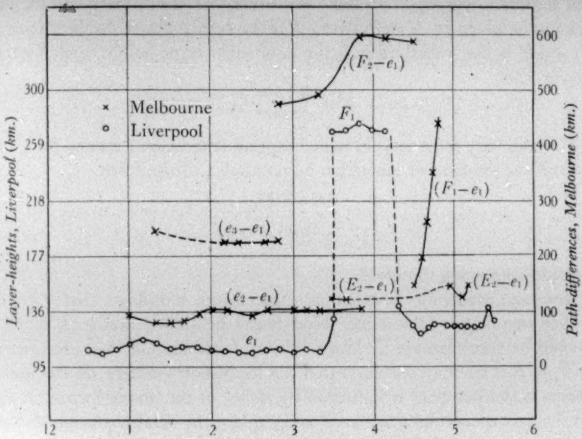


Figure 5.

The agreement between the equivalent heights of both the *F* and *E* regions measured at Melbourne and Liverpool is very close. Thus between 1 a.m. and 3 a.m. the average heights of the *e* layer deduced from (*a*) Liverpool observation, (*b*) the difference (e_3-e_1) and (*c*) the difference (e_2-e_1) are (*a*) 108 km., (*b*) 109 km., and (*c*) 109 km. respectively.

At 4.20 a.m. reflection at Liverpool occurs from the *E* region, but evidence of the presence of the *e* region is obtained for the shallow-range e_1 ray in Melbourne until 5.10 a.m., when the absorption of the waves became too great to permit of further observations.

§ 5. CONCLUSIONS

It has been remarked that interference fringes are always observable in Melbourne. The presence of fringes has previously been noted by one of us$^{(5)}$ on a longer wave-length (351 m.).

It is concluded that on broadcasting frequencies, at distances of the order 700 km., there are normally present at least two sky waves. The two principal sky waves are usually of comparable intensity, so that a large part of the fading experienced in such circumstances must be due to interference between these waves.

One of the most striking features of the results is the general uniformity of both the E-layer and the F-layer equivalent heights over the 700 km. between Melbourne and Sydney, at times when these layers are comparatively steady. Now the equivalent height of that part of the path of a ray which lies in an ionized layer depends on cos i, so that if the path in the layer is appreciable in comparison with the total path of the ray, we should observe appreciable differences between the layer-heights recorded simultaneously in Melbourne and Liverpool. For example, if the gradient of ionization be given by $N=\alpha h$, where h is the height above the beginning of the layer$^{(1)}$, and h' is the contribution to the equivalent layer-height made by the path in the layer, then

h
h'

$$h'=\frac{mp^2\cos^2 i}{2\pi\alpha e^2},$$

so that h' would vary some tenfold in our experiments as cos i ranged from 1 to 0·3.

Again, if the gradient of ionization be parabolic, of the form

$$N=\beta^2 h^2,$$

then

$$h'=\frac{\sqrt{(\pi m)}\,.\,p\cos i}{4\beta e},$$

and h' would vary some threefold.

For constant gradients of either of these forms it follows that h' cannot be greater than one or two kilometres. Now it can be shown that such small values for h' would be accompanied by lower values of attenuation than are observed in practice$^{(6)}$, so that we may conclude that the ionization gradient on the under side of the layer is sharper than is indicated by either of the above forms. It appears however that an exponential gradient$^{(6)}$ will explain the small variation of h' with i.

It may be concluded that, whatever be the ionization gradient, the E and F layers are both normally comparatively uniform in characteristics over the 700 km. path between Sydney and Melbourne.

During the occurrence of the dip in the F layer equivalent height which occurs about 3 a.m., however, it appears that there is a definite difference between the time of occurrence of the minimum heights over Sydney and Melbourne. The time lag in Melbourne is greater than could be accounted for directly by solar rotation.

Finally there is clear evidence that the average ionization density in the E layer over the whole 700 km. path of observation varies considerably from night to night. The significance of these variations will be discussed in another paper.

§ 6. ACKNOWLEDGMENTS

This work has been carried out as part of the programme of the Australian Radio Research Board. The authors wish to express their thanks to Mr W. J. Wark, M.Sc., for assistance in the experimental work, to Prof. T. H. Laby, F.R.S., for the provision of facilities in the Natural Philosophy Department of the University of Melbourne, to Mr A. H. Mutton, B.E., who was responsible for the operation of the emitter, and finally to Prof. J. P. V. Madsen for his continued encouragement and advice throughout the progress of the investigation.

REFERENCES

(1) Appleton. *Proc. phys. Soc.* **41**, 43 (1928).
(2) Appleton. *J. Instn elect. Engrs*, **71**, 642 (1932) (for bibliography).
(3) Appleton and Barnett. *Proc. roy. Soc.* A, **109**, 621 (1925).
(4) Appleton and Builder. *Proc. phys. Soc.* **45**, 208 (1933).
(5) Cherry. *Bull. Counc. sci. industr. Res. Austr.* No. 63 (1932).
(6) Martyn. *Proc. phys. Soc.*, p. 323 of this volume.
(7) Ratcliffe and White. *Proc. phys. Soc.* **46**, 107 (1934).
(8) Schafer and Goodall. *Proc. Inst. Radio Engrs*, N.Y., **19**, 1434 (1931).

620.152.272 546.815:620.172

AN INTERFERENCE EXTENSOMETER AND SOME OBSERVATIONS ON THE ELASTICITY OF LEAD

By BRUCE CHALMERS, B.Sc., Ph.D., Lecturer in Mathematics and Physics, The Sir John Cass Technical Institute.

Received October 19, 1934. *Read December* 7, 1934.

ABSTRACT. A description is given of an extensometer in which interference fringes are used to measure elastic and plastic extensions of specimens of length about 3 cm. to an accuracy of about 3×10^{-7} cm. Experiments on specimens of lead are described, and the following are the principal results obtained. (i) A specimen that has not recently been severely strained has a definite range in which Hooke's law is obeyed within experimental limits, and a definite elastic limit. (ii) When the specimen has recently been severely strained, a new type of closed elastic-hysteresis loop is obtained. (iii) The elastic after-effect was investigated, and it was found that when the stress is below the elastic limit, the whole of the observed effect can be accounted for thermodynamically. (iv) The true plastic after-effect (creep) was found to commence when the elastic limit is exceeded.

---

§ 1. INTRODUCTION

It is the object of the experimental study of elasticity to express the amount of a strain as a function of the stress producing it. To a first approximation, such a relationship is supplied by Hooke's law, the law of linear variation of strain with stress. More refined investigations have not only shown that deviations from Hooke's law occur, but have indicated that the stress is not the only variable on which the strain depends. More explicitly, it is found that the strain may vary with time as well as with stress, and that even when this time-variation has been allowed to reach completion, the strain is not necessarily a single-valued function of the stress, but may depend upon previous values of it. The stress-and-strain relations for a single specimen of a particular material are thus of great complexity; and when the material is a metal a further complication is introduced by the fact that the elastic properties of a metal depend upon the thermal and mechanical treatment to which the metal has been subjected, as well as on the purity of the metal.

It is evident, therefore, that in order to arrive at fundamental knowledge of the mechanism underlying the real elastic properties of metals, investigations must be made on specimens in some standard state of mechanical and thermal treatment. Now the effect of such treatment is to modify the size and disposition of the constituent crystals of the specimen. The standard state is therefore the state in which the specimen consists of a single crystal. The experimental difficulty of preparing very large single metal crystals sets a limit to the size of specimen readily obtainable, and the experimental method designed for the investigation of the elastic properties of single metal crystals should therefore be suitable for small specimens.

The experimental study of elasticity has been severely handicapped in the past by the fact that torsional and flexural strains are easier to measure than extensional strains. In both the former types of strain, the stress, and therefore also the strain, varies throughout the cross-section of the specimen. This, however, is not the case with extensional stresses which give homogeneous strains.

The requirement of an appropriate apparatus designed for investigations on the elastic properties of single crystals of metals is, therefore, the measurement of the extensions produced by known tensions acting on short specimens of the metal. The apparatus described below fulfils this requirement, and although experiments on single crystals have not yet been made, the results obtained in experiments on polycrystalline lead seem to be of sufficient interest to justify separate publication. In particular, a type of elastic hysteresis loop which has not previously been noticed has been observed to occur in certain conditions, while in other circumstances lead has a definite elastic range in which Hooke's law is obeyed.

§ 2. EXPERIMENTAL METHOD

The apparatus consists essentially of two parts, of which the respective functions are to apply a known stress to the specimen and to measure the strain produced. The main frame of the apparatus consists of two vertical brass plates, each about 25 cm. square, held parallel to each other at a distance apart of about 8 cm. by horizontal brass bars screwed to the plates. The disposition of the cross-bars is shown in section in figure 1 (*A*, *F*, *Z*, *Z*).

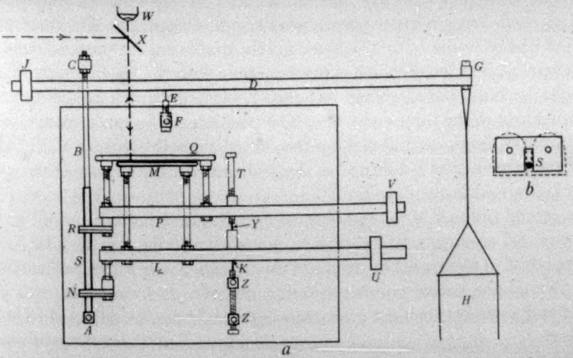


Figure 1.

Application and measurement of the stress. The specimen *S*, figure 1 *a*, is soldered at its lower end to the centre of the cross-bar *A*, and its upper end is soldered to the end of the brass rod *B*, so that *S* and *B* are collinear. The upper end of *B* is fixed to a cross bar *C* through which an upward force is applied to *B* and *S*. Near each end of *C*, $2\frac{1}{2}$ cm. from *B*, is a screw held vertically by a screw thread in *C* and a lock nut, and from the lower end of each of these screws a steel gramophone needle point projects. These points rest in punch marks on the two bars D, which are rigidly connected to form the lever through which the stress is applied. The bars D pivot on the points of two gramophone needles E carried by screws passing through the cross bar F of the main frame.

The ends G of the bars D are connected by a cross-bar, and this cross-bar supports a glass tube H of length 25 cm. and diameter 6 cm., closed at its lower end and open at the top. Water can be introduced into or removed from the tube H by means of a two-way syphon (not shown in diagram).

The stress applied to the specimen is directly proportional to the amount of water in the tube H in excess of that required to balance the counterpoise weight J which is supported by the other end of the lever DD. Since the syphon tube dips into the water, allowance must be made for the downward force which it applies; this force is equal to the weight of the water displaced by the tube. Hence the level of the water-surface in H indicates the stress applied. The water-level is determined by means of a pointer which is moved vertically so as just to touch the water-surface. The pointer is supported by a glass rod which moves along a scale. The tube H is calibrated by introducing known amounts of water and observing the change of water-level.

Measurement of the strain. In order to determine the strain of the specimen corresponding to any stress, it is necessary to measure changes in the distance between two points on the specimen. These measurements were made by an optical interference method as follows. Two screws K, 4 cm. apart as measured along the bars Z, Z, each carry a gramophone needle, point upwards. On these points is supported a brass frame L, to which are rigidly attached a flat piece of glass M, and an arrangement N through which contact is made with the specimen S. N consists of two similar brass plates, shown in figure 1, between which are held two halves of a razor-blade of the three-hole type. The position of the razor-blades, which lie in a horizontal plane, is indicated by the dotted lines in figure 1 *b*. The distance apart of the razor-blades is adjusted so that they just cut the surface of the specimen when it is inserted in the position S.

The second frame P of the interferometer, figure 1 *a*, holds a second glass plate Q and a second arrangement R similar to that shown in figure 1 *b*. The frame P is supported on L at two points by means of two screws T, the actual points of support being gramophone needle points projecting downwards from the screws T. The distance and angle between the glass plates Q and M can be adjusted by means of the screws T.

The frames L and P are each counterbalanced by the adjustable weights U and V, so that no stress is applied to the specimen S through R or N. The interferometer is illuminated from above by means of light from a mercury arc passing through a water cell and a monochromatic green filter, and reflected downwards from the glass slip X. The fringes produced by the interference between the light reflected upwards from the lower surface of Q and from the upper surface of M are viewed through the microscope W. The direction and spacing of the fringes, which are the loci of points of equal separation of the plates M and Q, can be adjusted by means of the screws T. This adjustment is made so that the fringes form a series of parallel lines whose direction is perpendicular to the plane of figure 1*a*, i.e., parallel to the line TT of the fulcrum of the interferometer unit. When the distance RN alters, the fringes move in a direction perpendicular to their length. The plates Q and M were pieces of good plate glass mirror, and it was found that the fringes obtained when there was no silvering on either plate were sufficiently sharp and sufficiently straight for satisfactory measurements to be made.

Precautions taken to avoid disturbances. In order to avoid disturbances due to vibrations of the bench on which the apparatus stood, the two frames P and L of the interferometer were held together by two steel springs in the plane of the two screws T. The springs, being in the plane of the fulcrum about which the upper frame moves on the lower one, exert no force on the specimen through the interferometer. A second precaution to avoid vibrations was to fix the syphon arrangements and the apparatus for measuring the water-level in H on a different bench from the one on which the apparatus was placed. The apparatus itself was supported in a water bath on pads of rubber sponge which almost entirely prevented any effect from slight vibrations of the bench. In spite of these precautions it was occasionally found that vibrations occurred which altered the position of the fringes. This did not occur often, and did not cause serious difficulty in the experiments. Smaller vibrations made the fringes appear to broaden for the duration of the vibrations, but did not permanently alter the position of the fringes.

The surface of the water in the water bath was just below the level of the lower interferometer mirror, the specimen, contact pieces and supporting screws being under water.

Design of apparatus. A certain degree of compromise has been necessary in designing the apparatus, since the conditions necessary for the elimination of vibration appear to be incompatible with strict geometrical considerations. It must be realized, however, that the components of the instrument only turn through very small angles, seldom exceeding 10^{-4} radian.

The particular instance in which this applies is that of the lower interferometer frame L. The point of support K should be in the same horizontal plane as the point of contact N, as the point N must move in a direction perpendicular to the line NK. The reason why this condition was not observed was that it was considered essential that the upper frame P should rest directly on the lower frame L and not on a second cross-bar, in order to obtain consistent readings and to avoid vibrations. The point Y should therefore not be allowed to move horizontally when L rotates, so Y must not be far above the point of support K. Since the upper frame P moves considerably more than the lower frame L, it is desirable that Y and R should be in the same horizontal plane. Thus K must be higher than N. Since N is near the lower end of the specimen, its movement during an experiment is very small; hence the frame L turns through a very small angle (not exceeding 2×10^{-5} radian). It follows that the horizontal reaction on N during an experiment is very small, and that the horizontal movement of N will be small compared with the relative vertical motion of R and N. It follows that the error in the results due to non-observance of the strict geometrical considerations of design will be negligible. The results obtained indicate that this is the case; see § 3.

Experimental procedure. The experimental procedure depended upon the nature of the variation under investigation, and in general consisted in first adjusting the screws T so that the spacing of the fringes seen in the microscope was suitable, then allowing the temperature of the apparatus to become steady or, more usually, uniformly rising. The movement of the fringes caused by thermal expansion of the specimen and apparatus owing to this slow rise of temperature was allowed for in the results.

Movements of the fringes due to changes of water-level in H were investigated by changing this water-level by equal steps and observing the movement of the fringes during and after the change of stress. The movements of the fringes were measured and expressed in terms of the fringe-spacing as a unit.

Accuracy of the method. The length of the specimen between R and N was about 3 cm., and the change of this length which will cause a displacement of the fringe system by one fringe space can be estimated as follows. When the fringe system is displaced by one unit, the change in the distance between Q and M at the point of observation is one-half wave-length of the light used. The mercury green line 5460 A. was used in all the experiments, so that the change of distance between Q and M for one fringe shift is $\frac{1}{2} \times 5460$ A. or $2{\cdot}73 \times 10^{-5}$ cm. This is the change of separation of Q and M at the point on which the microscope is focused. The increase Δl of the length l of the specimen is greater than this because the specimen is farther than the point of observation from the fulcrum; hence the change in separation of QM must be multiplied by a factor, which was found to be 1·35. The change of length of the specimen corresponding to a shift of one fringe is $2{\cdot}73 \times 1{\cdot}35 \times 10^{-5}$ cm. or $3{\cdot}68 \times 10^{-5}$ cm., so that $\Delta l/l$, the extension per unit length per fringe $= 3{\cdot}68/3{\cdot}00 \times 10^5 = 1{\cdot}23 \times 10^{-5}$ approximately.

The microscope used for observing the fringes contained an eye-piece scale. By adjusting the fringes to be about ten scale-divisions apart, and reading the position of a fringe and the distance between two fringes to one-tenth of a scale division, movements of the fringes could be determined to 0·01 of a fringe. It was usually found advantageous to observe two or three fringes. A movement of 0·01 fringe represents a value of $\Delta l/l$ equal to $1{\cdot}23 \times 10^{-7}$.

In the experiments described below, the position of the fulcrum E was such that a movement of one fringe was obtained when the water-level in H was changed by about 1 cm. The position of the water-level could be read to about 0·1 mm., so that the stress was measured to about the same accuracy as the strain. The tube H was calibrated by introducing known volumes of water from a burette, and observing the corresponding rise in water-level.

When observations of $\Delta l/l$ are made to an accuracy of the order of 10^{-7} cm./cm., the temperature of the specimen assumes considerable importance. The coefficient of expansion of lead is $2{\cdot}76 \times 10^{-5}$ cm./cm. ° C., so that an increase of length of 10^{-7} cm./cm. is produced by a rise of temperature of about 0·003° C. A correction for change of temperature during the course of a series of observations was made when necessary from a calibration curve giving the movement of the fringes which occurred with change of temperature, when other things were equal. The water bath was kept vigorously stirred and its temperature was measured to 0·01° C. The rate of change of temperature of the water bath was always small and steady, and never exceeded 0·5° C. per hour. The results quoted in § 3 show that the accuracy obtained was of the order anticipated.

§ 3. EXPERIMENTAL RESULTS

The experimental work has consisted in the investigation of the variation of the distance between two points on a particular specimen of lead with the applied tension and with time. The specimen was a lead wire, 3 mm. in diameter, containing less than 0·001 per cent impurity, to which no special heat treatment had been applied. The results obtained can be described in general as follows. The variation of the length of the specimen with tension and time depends upon the history of the specimen, and in particular on whether the specimen has recently been subjected to a stress above a certain critical value. The results will therefore be considered in two sections, namely those obtained on an unstrained specimen, and those obtained after the specimen has been strained.

The results obtained with the unstrained specimen are comparatively simple, Hooke's law being obeyed within a small but definite stress-range. When this range is exceeded, the behaviour of the specimen deviates from Hooke's law in the usual way. These results are described in detail in the next paragraph. If, however, the specimen has previously been subjected to a stress much larger than the stress at which deviation from Hooke's law occurs, and if it has been allowed sufficient time for its length to become constant, then a type of elastic hysteresis manifests itself. The formation of the hysteresis loop, which is of a type not previously described, is discussed at the end of this section. The experimental study of these effects is complicated by the presence of elastic after-effects.

Elasticity of the unstrained specimen. Two typical stress-and-strain curves obtained with an unstrained specimen are shown in figure 2, in which graph (*a*) shows a case in which the elastic limit has not been exceeded, while graph (*b*) shows a case in which it has. The points on (*a*) lie on a straight line, their greatest deviation not exceeding 5×10^{-7} cm./cm. The points for increasing stress are represented as crosses, while those corresponding to decreasing stress are shown as circles. The two sets of points lie on the same line, and there is no observable permanent set of the specimen. The points shown were obtained by increasing or decreasing the stress by measured amounts, and waiting for three minutes after each change of stress before taking the reading of the fringes. This was necessary because the specimen does not attain the final length corresponding to any stress for a few minutes, eventually reaching a definite length. This elastic after-effect was observed whenever the stress was changed, and will be further described below.

The curve (*b*), figure 2, shows that there is a limit to the region in which Hooke's law is obeyed. This limit occurs when the strain is just above 3×10^{-5} cm./cm. The critical stress is 9·0 kgm./cm.^{2} This point is also the lower limit of strains which result in permanent set observable by the present method. A third characteristic of this critical strain is that when it is exceeded a change takes place in the type of elastic after-effect obtained.

The method of investigation of the after-effect was to change the tension on the specimen by definite steps, usually either about 1500 gm./cm.^{2} or 3000 gm./cm.^{2} corresponding to changes of water-level in the tube H, figure 1 (*a*), of 1 cm. and 2 cm. respectively. These changes of tension were made in the direction of either increasing or decreasing stress, and after each change the movement of the fringes was observed until it was less than 0·01 fringe per minute, which corresponds to a change of length of the specimen of about 10^{-7} cm./cm. per minute.

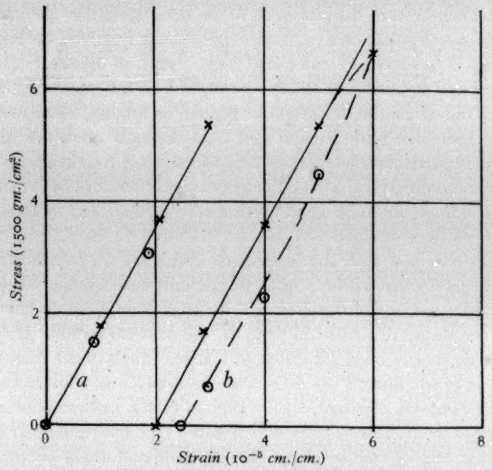


Figure 2.

The type of curve obtained when the tension did not exceed the elastic limit was always that shown in figure 3 (*a*) in which several curves, selected at random, for these changes of stress are shown. It is found that similar curves are obtained with increasing and decreasing stresses, and that so long as the elastic limit is not exceeded the curve depends only on the change of tension and not on the tension itself.

Some uncertainty as to the time zero of these curves is inevitable, for the stress cannot be increased or decreased suddenly without causing vibrations that prevent measurements from being made. With the method of changing the stress adopted, i.e. syphoning water into and out of the tube, the change of stress takes at least 3 seconds and usually considerably more, and in general the zero of the curves could not be fixed to within about 10 seconds. This uncertainty will be referred to again in the discussion of results.

Curves of the type shown were always obtained unless the elastic limit was exceeded. When the stress applied was above the elastic limit, the after-effect no longer ceased after a few minutes, but continued for a much longer time in a manner similar to that described in the next section.

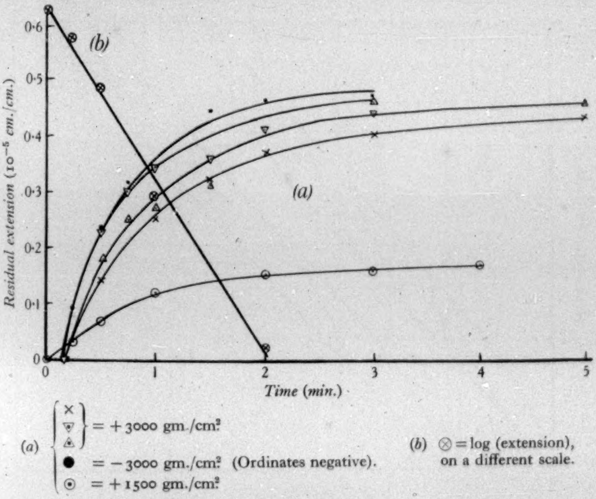


Figure 3.

Elastic hysteresis of the strained specimen. If the stress is increased by steps from zero up to a maximum and then reduced to zero, and if a reading of the length is taken for each stress after sufficient time has elapsed for the lengths to have become constant, a closed loop denoting the existence of hysteresis is obtained. A series of such curves corresponding to different maximum stresses is shown in figure 4 (*a*).

Each loop consists of two curved parts and two straight lines. The curved parts are similar for all the loops of the series, and the same two parallel straight lines form the straight parts of all the loops. It follows that the width of the loop, i.e. the difference in the lengths which the specimen has when the stress is increasing and when it is decreasing, is independent of the maximum stress reached, provided that the latter is less than the elastic limit. The width of the loop and the slope of the straight lines vary with temperature and with the crystalline state of the metal; for instance they depend on whether the elastic limit has recently been exceeded. The width of the loop obtained with the same specimen varied, in the experiments described, between $0{\cdot}5 \times 10^{-5}$ cm./cm. and $2{\cdot}5 \times 10^{-5}$ cm./cm. according to the amount of strain to which the specimen had been subjected.

A further observation regarding the hysteresis loop is that the lower extremity of the loop can be produced at any stress-value by increasing the tension from that point. A typical example demonstrating this observation is given in figure 4 (*b*). The stress was increased from *A* to *B*, then decreased from *B* to *C*, increased from *C* to *D* and finally reduced from *D* to *A*. The point *C* behaved in a manner similar to the point *A*. Hence the same loop is obtained whether the lower reversal is made at zero stress or at some other stress. A further point of importance is that, although in these experiments stress is only being applied in one sense, i.e. tension and not compression, the loop is closed, and can be repeated a number of times without observable change.

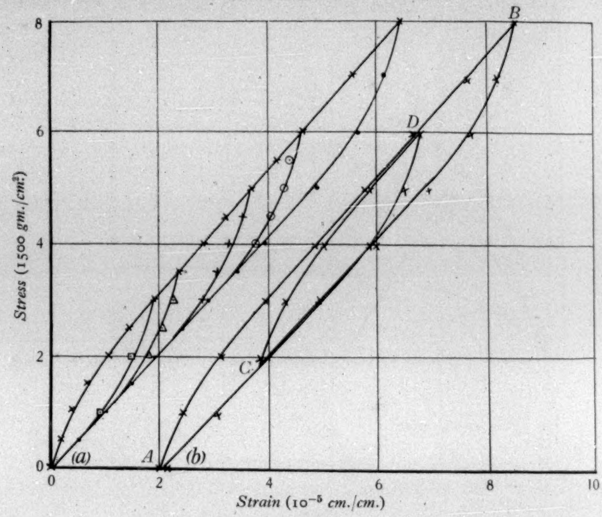


Figure 4.

The elastic limit. The constancy of the slope of the straight-line parts of the hysteresis loops is shown in figure 5, in which the change of length, in units of 10^{-7} cm./cm., corresponding to a change of stress of 1500 gm./cm.², is plotted against the total stress. The points marked by crosses indicate increasing stress, and the circles decreasing stress. The points lie very close to a straight line, with the exception of two points at the beginning of each set which correspond to the curvature which gives rise to the hysteresis loop. Apart from these points, the maximum deviation from a horizontal straight line is three units, indicating a maximum difference of slope of the stress-and-strain line of 3 per cent from the mean. There is, however, no systematic deviation of the points from the horizontal line, and it follows that there is no change of slope above the limits of the experimental error. It is safe to say that the slope does not change by more than 1 per cent over the range of this curve.

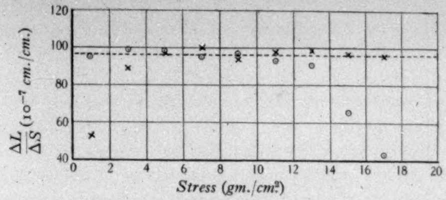


Figure 5.

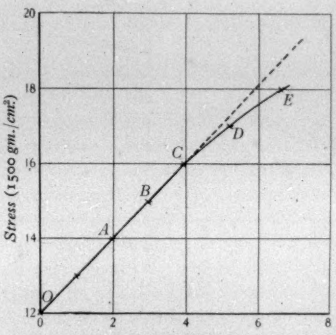


Figure 6.

Figure 6, however, shows that at a greater stress an abrupt departure from the straight line takes place. *OABC* represents the straight line part of the curve, and beyond *C* an obvious deviation occurs. Curves showing the elastic after-effect corresponding to increases of stress from *A* to *B*, *B* to *C*, *C* to *D* and *D* to *E* are given in figure 7. It is clear that the point *C* marks the transition between the two kinds of after-effect curve, as well as the end of the straight line. The points above *C* are not as definite as the points below *C*, because the after-effect curve shows no definite final value once *C* has been passed. It follows that the point *C* marks a definite change in the response of the metal to stress, and is the transition point between the elastic and the plastic parts of the curve. The point *C* is therefore taken as the elastic limit. It will be shown in § 4 that the after-effect occurring before this point is due to thermal and not to plastic effects. The amount of stress corresponding to the elastic limit depends upon the history of the specimen, and so cannot be given a definite value. When the greatest strain to which the specimen has been subjected is comparatively small, i.e. not very much above the primitive elastic limit, the elastic limit of the strained specimen is low; when the maximum stress is high, the elastic limit is also high.

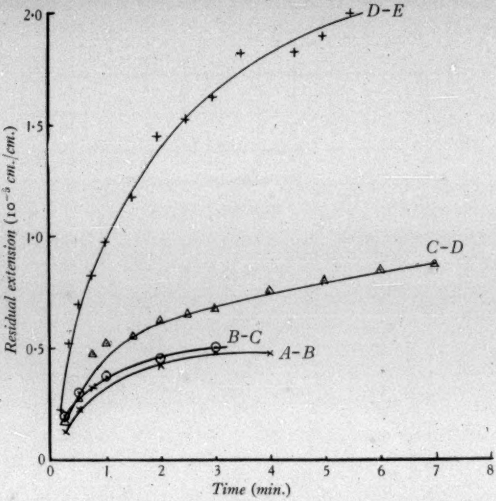


Figure 7.

§ 4. DISCUSSION OF THE RESULTS

The experimental study of elastic hysteresis and the allied effects has been pursued in the past chiefly by the use of torsional stresses and strains(3, 5, 8). Such work, although of the greatest value from the practical point of view, cannot be made to give very much information regarding the fundamental relations between stress and strain. The effect of a torsional stress on a material which does not obey Hooke's law for the whole range of stress applied must be to set up a very complicated set of internal stresses and strains. The observed torsional deformation would be some function of the stress and of the deviation from Hooke's law. If, in addition, the stress acting on some part of the specimen, namely the outer shell, is above the limit of approximate adherence to Hooke's law, while the rest of the specimen is under stresses varying from zero to this limit, effects resembling the production of a hysteresis loop may result, the energy-loss in the loop being produced in the part of the material which is in the plastic state.

Much of the work done on elastic hysteresis in which a homogeneous stress (e.g. tension) is used is valueless from the present point of view owing to insufficient accuracy. Ewing(1) performed experiments on long wires with an accuracy of 10^{-7}, but his experiment only established the existence of a hysteresis loop without investigating its shape. Gough, Hanson and Wright(2) used tensile tests for part of their work, making their measurements with a Marten's extensometer, the limit of their accuracy being 2×10^{-6}.

In these cases, as in others in which the apparatus is constructed on a fairly large scale, sufficient temperature-control for an accuracy approaching 10^{-7} cannot be obtained.

The degree of accuracy obtainable with the present apparatus is necessary in order to establish the existence of a definite elastic range in the unstrained specimen. The elastic range is regarded as being definite because there appears to be a definite point at which three deviations from perfect elasticity first become discernible—departure from the straight line, permanent set, and true plastic flow. These deviations increase rapidly, and roughly in proportion to the amount by which the stress exceeds the elastic limit, when this limit has been passed. These observations indicate that a higher degree of accuracy would not indefinitely reduce the elastic limit.

After the specimen has been strained and allowed to recover from the immediate effects of the strain (which take the form of a continuous reduction of length towards a final value to which it approaches asymptotically and is sensibly equal after a few hours) its behaviour is quite different from the elastic behaviour discussed above. The slope of the stress-and-strain line at the origin is roughly equal to that of the straight line of the unstrained specimen, but further increase of stress reduces the slope of the line, so that there is less stress per unit strain, and the slope of the straight line which is eventually reached is between one-half and two-thirds of that of the primitive elastic line. Hence the general effect of applying a large stress to the specimen is to reduce its effective elastic modulus, bring it to a condition in which hysteresis loops can be obtained, and replace the low elastic limit of the unstrained specimen by an elastic-hysteresis limit which is much greater (the limiting stress being five or ten times as great) and shares all the properties of the actual elastic limit.

The type of hysteresis loop obtained in the investigations cited above, and discussed by Tomlinson(9) and by Prandtl(6), is that shown in figure 8. Such a loop depends upon the deviation of the line of increasing stress $ABCD$ from a straight line at C. It requires, then, that the stress applied, at any rate to part of the specimen, should exceed the elastic limit. The loop is a result of permanent set of the specimen occurring at the extremes of stress, and can only be closed if equal stresses of opposite signs are applied alternately. Since it would presumably not exist if the reversal of the stress were made at the point C, it depends upon the material entering the plastic or overstrained state, and the name ***plastic hysteresis*** may be suggested for this type of loop.

The formation of the hysteresis loop observed in the present experiments differs fundamentally from the plastic hysteresis loop referred to above. An elastic hysteresis loop, such as that shown in figure 4, depends upon the fact that the stress-and-strain line is curved either when the stress is increased from a minimum value or decreased from a maximum value, irrespective of what these values are, provided that they do not exceed the elastic limit, i.e. that the stress is not sufficient to produce a sensible deviation from the straight-line part of the stress-and-strain curve, or to cause a true plastic after-effect exceeding 10^{-7} cm./cm. per minute after 5 minutes.

In a recent paper on the bending of marble, Lord Rayleigh(7) gives diagrams of the hysteresis loops he obtained by applying flexural stresses to specimens of marble that had been heated. The diagrams show loops very similar to those described in the present paper, and appear to be elastic rather than plastic in origin. Similar elastic hysteresis loops are given by Hanson(4) for zinc single crystals by bending. The difference between these loops and the plastic loops is not discussed in either of the papers cited.

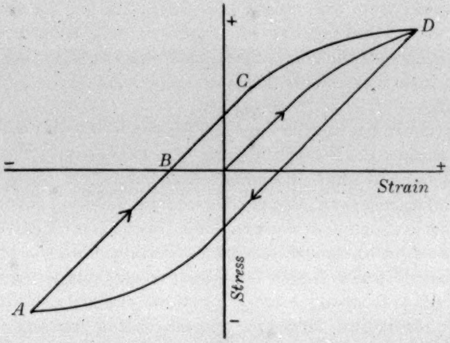


Figure 8.

Before discussing the significance of these loops it is necessary to enquire into the possibility of their existence being only apparent, and due to some defect in the apparatus. Various experiments have been made with a view to eliminating this possibility. Any cause which would prevent the upper plate of the interferometer from moving freely, so that it lagged behind its true position, would cause a loop such as was found to appear. Two possible ways in which this might happen suggest themselves, the first being that some friction might prevent free movement of the upper plate, and that the upper interferometer mirror might not move in proportion to the movement of *R* (figure 1) until *R* had moved by a certain amount. This would cause the length to appear too small when it was increasing and too large when decreasing. A similar effect on the lower plate would give the reverse effect, and a loop of apparently negative area would result.

The effect of friction is usually to prevent all movement until a limiting force is reached, when movement suddenly starts; this is not what happens with the hysteresis loops, since the ends of the loops are curved and not flat. Further, if such a frictional force existed, there would be a range of positions in which the upper interferometer plate would be in equilibrium between the frictional force and the force acting on it at its point of contact with the specimen. It is found, however, that if the plates are set in vibration by lightly tapping the bench on which the apparatus stands, after the broadening of the fringes due to vibration has disappeared, the fringes are in the same position as before, to an accuracy of 0·01 of a fringe. This indicates that there is only one position of equilibrium, and that friction does not affect the position of the fringes to more than 0·01 of a fringe. It was also found that if the plate *R* was moved by amounts corresponding to a movement of one or two fringes by pressure of the finger on *R*, *T* or *V*, it returned to its original position when the pressure was removed.

The second way in which spurious loops might appear is that the contact at *R* might not be effective, and might be different when the specimen was moving upwards and when it was moving downwards. On one or two occasions when the apparatus had been set up so that contact at *R* was definitely unsatisfactory, it was impossible to get the fringes steady, owing to chance vibrations of the upper plate. It follows that when the fringes are steady contact is satisfactory. A further test was made as follows. Normally the upper interferometer plate was balanced and applied no force to the specimen; stress-and-strain curves were taken after altering the balancing so that *R* applied first a slight upward force on the specimen at the point of contact, and then a downward force. Identical loops were obtained in the various cases.

A further consideration is that the appearance of the loops did not depend upon the particular setting of the apparatus, for they were not altered by taking the specimen out, taking the apparatus to pieces and putting it together again, and replacing the specimen.

The most positive consideration, however, is that it is possible to get either the straight line or the hysteresis loop with the same specimen without removing it from the apparatus or making any other alteration except applying a larger stress to the specimen. In no case did a specimen which normally gave a straight line give a loop, and a specimen normally showing a loop was never observed to give a straight line. It must be concluded that it is not the instrument but the specimen which causes the loop to be observed, and that the hysteresis loop is an actual property of the specimen.

It is next necessary to discuss the significance of the elastic hysteresis loop. In the first place, the magnitude of the loop can be expressed as the ratio of the work transformed into heat during the cycle to work done on the specimen during its extension. In the series of loops given in figure 4 (*a*) this ratio varies from 0·21 for the largest loop to 0·33 for the smallest loop.

It follows that if a specimen of lead which has previously been strained is set in vibration, up to about one-third of the energy of the specimen in a single vibration may be lost if the amplitude is small. This conclusion will account for the fact that elastic vibrations in lead die away very quickly, and may account for the observation of Waller(10) that lead does not share the property of most other metals of being set into acoustic vibration by contact with solid carbon dioxide. This observation must be distinguished from the decay of torsional vibrations of wires, in which plastic losses are probably concerned.

The general cause of the loss of energy involved in the hysteresis loops must be some kind of internal friction that causes a difference in the equilibrium between stress and strain when the stress is increasing or decreasing. It would be unprofitable, however, to attempt to picture the mechanism involved until the way in which the effect varies with crystalline arrangement has been investigated much more fully and, in particular, until it has been established whether the hysteresis loop can be obtained with a specimen consisting of a single crystal.

Although the elastic after-effect described above in connexion with figure 3 superficially resembles the true plastic after-extension, it can be shown that it may be due to thermal changes and not to plastic flow. It was pointed out by Joule that if a substance expands when heated, then if that substance is subjected to pressure its temperature will rise. Conversely, if a tension be applied the temperature will fall. Calculation shows that the thermal expansion due to the rise of temperature required to bring the specimen back to the temperature of its surroundings, after a fall in temperature due to the application of a tension, is of the right magnitude to account for the observed after-extension. A reduction of the tension gives a corresponding effect with the sign changed. The calculation only applies while the stress applied is less than that required to cause deviations from the straight line *OAB*, figure 6.

The fourth thermodynamic relation of Maxwell states that

$$\left(\frac{dH}{dp}\right)_T = -v\alpha T,$$

where dH is the heat entry during an isothermal change due to a change of pressure dp of a volume v at a temperature T, the coefficient of expansion being α. The isothermal heat entry dH is also the heat that will enter during the return to the original temperature if the first change is adiabatic, and that is approximately the condition of the experiment. Hence the fall in temperature of the specimen during an adiabatic extension can be calculated, and so the subsequent extension due to rise in temperature can also be determined.

$$dH = -v\alpha T dp,$$

or

$$\Delta H = -v\alpha T \Delta p,$$

where Δp is the change in tension in dynes/cm.^{2},

$$H = -\alpha T \text{ erg/cm.}^3 \text{ per dyne/cm.}^2,$$

or

$$H = -\alpha TD/J \text{ cal./gm. per dyne/cm.}^2,$$

where J is Joule's Equivalent and D is the density. The corresponding change in temperature ΔT is given by

$$\Delta T = -\alpha TDg/JS \text{ deg. per gm./cm.}^2,$$

where S is the specific heat.

Hence the subsequent increase ΔL of unit length is

$$\Delta L = -\alpha^2 TDg/JS \text{ cm. per gm./cm}^2.$$

Evaluating this for lead we get

$$\Delta L = 1\cdot937 \times 10^{-9}.$$

When the increase of tension is 3000 gm./cm.^{2},

$$\begin{aligned}\Delta L &= 3000 \times 1\cdot937 \times 10^{-9}\\ &= 5\cdot811 \times 10^{-6} \text{ cm./cm.},\end{aligned}$$

which is about the extension required to give a shift of one half-fringe.

It was always found that a change of tension of 3000 gm./cm.^{2} within the elastic limit gave an after-effect of less than 0·5 of a fringe, see curve *A*, figure 4. For greater changes of tension greater after-effects were found, but never greater than the value calculated as above. There are two reasons why the observed after-effect should be less than that calculated. The first is that the change is not strictly adiabatic, as it takes a definite time to apply the stress; hence the fall in temperature T is not as large as is found in the calculation. Secondly, observations of the positions of the fringes cannot be taken immediately after the stress is applied, so the full extent of the after-effect is not observed.

The slight variation in the shape of the curves *A*, *B*, *C*, *D*, figure 4, is probably due to the fact that the rate of loss of heat from the surface of the specimen depends upon the speed of circulation of the water surrounding it. This varied according to the speed of the stirring-motor. Hence the shape, but not the final extension, may vary.

The variation of temperature of a cylinder with time, if the cylinder loses heat radially to its surroundings, can be derived from a Bessel-Fourier function. The variation of mean temperature with time can be shown to follow an exponential law. When one of the curves of figure 3 is plotted logarithmically (i.e. when $\log\{(L-L_\infty)/(L_0-L_\infty)\}$ is plotted against t, where L is the length at time t, L_0 the length when $t=0$, and L_∞ the final length) the curve of figure 3 (*b*) is obtained. The fact that this curve is straight shows that the law of the thermal after-effect is exponential.

It is concluded, therefore, that within the elastic limit as indicated by the straight-line part of the stress-and-strain curve, the after-effect does not exceed the amount that can be accounted for on thermodynamic grounds. It appears probable that plastic flow commences at the elastic limit.

The curves of figure 7 show quite clearly that the after-extension completely changes its character when the point *C*, figure 6, has been passed, and that the rate of this plastic extension increases rapidly with increase of stress beyond that corresponding to the point *C*, figure 6. This applies in both the strained and the unstrained conditions.

Although Hooke's law is not obeyed with lead after straining, it is possible to express a value for Young's modulus for the specimen concerned; such a value refers to the slope of the straight-line portions of the stress-and-strain loop. Since the slope of these lines varies in some manner as yet unknown with the conditions that affect the crystalline arrangement of the metal, it is useless to attempt to evaluate Young's modulus to a high degree of accuracy. The values obtained are

$$Y = 1{\cdot}7 \times 10^{11} \text{ dyne/cm}^2 \text{ for a strained specimen}$$

and

$$Y = 2{\cdot}7 \times 10^{11} \text{ dyne/cm}^2 \text{ for an unstrained specimen.}$$

Although the actual stress corresponding to the elastic limit varied somewhat, it was found to be of the order of 30,000 gm./cm^2, and the limit of the elastic extension was in the neighbourhood of 2×10^{-4} cm./cm. for a strained specimen. The corresponding values were 9000 gm./cm^2 and 4×10^{-5} cm./cm. for the unstrained specimen.

REFERENCES

(1) Ewing. *Brit. Ass. Report*, p. 502 (1889).
(2) Gough, Hanson and Wright. *Philos. Trans.* A, **226**, 1 (1925).
(3) Guest and Lea. *Proc. roy. Soc.* A, **93**, 313 (1917).
(4) Hanson. *Phys. Rev.* **45**, 5, 331 (1934).
(5) Kimball and Lovell. *Phys. Rev.* **30**, 948 (1927).
(6) Prandtl. *Z. angew. Math. Mech.* **8**, 85 (1928).
(7) Rayleigh. *Proc. roy. Soc.* A, **144**, 266 (1934).
(8) Rowett. *Proc. roy. Soc.* A, **89**, 828 (1914).
(9) Tomlinson. *Phil. Mag.* **17**, 113, 634 (1934).
(10) Waller. *Proc. phys. Soc.* **46**, 116, 252 (1934).

DISCUSSION

Prof. Andrade. I understand that the strained rods recovered their unstrained condition, in which they showed no hysteresis, if kept at air temperature for a time. Is that so, and, if so for how long must they be kept? Did the author carry out any experiments on annealing strained rods, and what kind of strains did he employ? Were they simple extensions only? What were the magnitudes of the strains?

Comparison of figures 2 and 4 seems to show that the strained rods have a higher modulus than the unstrained to begin with, but a lower one when they reach their steady state, corresponding to the straight-line portion of the curve. Are figures 2 and 4 strictly comparable?

Has the author studied the flow when the rods are subjected to such a stress that the creep continues for hours? The thermodynamic cooling cannot, of course, be calculated on simple lines in this irreversible case, but must be exceedingly small. It would much interest me to be able to compare the flow at such relatively small stresses with that at much higher stresses which I measured many years ago. In the case of these small extensions no special device is needed to keep the stress constant as the rod extends.

As regards the propagation of sound, it looks as if an unstrained piece of lead should conduct sound quite well, as there is no hysteresis: the velocity will, of course, be that corresponding to adiabatic compressions and tensions, unless the frequency is very low. A strained piece of lead, however, exhibits hysteresis, and should dissipate energy and show marked attenuation.

Dr F. D. SMITH. It is difficult to find an entirely satisfactory definition of Young's modulus for a material like lead, which exhibits elastic hysteresis. We are not justified in calculating it from one part of the hysteresis loop, figure 4, rather than another. The velocity of propagation of elastic vibrations of large amplitude must be subject to a similar complexity. Certainly such vibrations will be rapidly distorted and attenuated.

Elastic vibrations of small amplitude travel well in lead, as Dr Wood and I have noticed in our experiments. It may be that the hysteresis observed by the author does not occur unless the strain is large.

Does the strain ever increase discontinuously? We know that intensity of magnetization sometimes increases discontinuously (Barkhausen effect). The extensometer must, of course, be undamped for this test.

It would be preferable to plot stress (cause) as abscissa and strain (effect) as ordinate.

Dr A. B. WOOD. The paper gives values of Young's modulus for lead ranging from $1{\cdot}7 \times 10^{11}$ dyne/cm.^{2} for a strained specimen to $2{\cdot}7 \times 10^{11}$ for an unstrained specimen. It is interesting to compare these values with those deduced from the velocity of sound in rolled lead sheet, as measured by a method recently described in the *Proceedings**. The velocity of sound in rolled sheet (presumably overstrained) was found to be $1{\cdot}56 \times 10^5$ cm./sec., which gives for the elastic modulus $E/(1-\sigma^2)$ a value of $2{\cdot}73 \times 10^{11}$ dyne/cm.^{2} Assuming a value $0{\cdot}446$ of Poisson's ratio σ, this gives $2{\cdot}35 \times 10^{11}$ dyne/cm.^{2} for Young's modulus E. It would be a very simple matter to determine by the velocity method the variation of elasticity due to work-hardening of lead sheet. This method has the advantage that the strains superposed on the permanent strain in making the measurement are always very small.

Dr MARY D. WALLER. I notice that the temperature is kept constant to within $0{\cdot}5°$ C. The temperature coefficient of variation of plasticity (possibly also that of elasticity) in the case of lead is probably considerable: would it not therefore be desirable to state the actual temperature of the experiments? The value of Young's modulus is higher than that usually quoted. Do wires of different diameters give constant results? The large diminution in Young's modulus as a result of straining is interesting and may be characteristic of this peculiarly plastic metal. It appears to be much greater than that of other metals for which data are available.

AUTHOR's reply. In reply to Prof. Andrade: the question of annealing at room temperature and higher temperatures has not yet been dealt with quantitatively, but my results show that a strained specimen reverts to the unstrained state when kept at room temperature for some days. The curves in figures 2 and 4 are strictly comparable, being plotted in the same units, and the conclusions drawn from a

* *Proc. phys. Soc.* **47**, 149 (1935).

comparison are valid. I have not yet obtained creep curves over long periods of time, but this should not present any particular difficulty, and it will be interesting to see whether these results agree with Prof. Andrade's empirical law for much faster rates of creep.

In connection with the transmission of sound by lead, mentioned by Prof. Andrade, Dr Wood and Dr Smith, it seems possible that the adiabatic stress-and-strain curve may not exhibit the hysteresis loop even when the isothermal stress-and-strain curve does so. This cannot easily be tested directly, but can perhaps be concluded from such considerations as those mentioned.

The importance of grain-size and the distribution of impurities in the material is probably considerable in connection with the elastic properties, and may render a comparison such as that suggested by Dr Wood of little value when the materials used have undergone very different treatments in their preparation.

No discontinuous increases in strain have been observed, and the experimental results give no indication even of a discontinuity of the slope of the stress-and-strain curve, except perhaps at the elastic limit.

In reply to Dr Waller: the experiments were all carried out at temperatures of about 15° C. Wires of different diameters have not been compared; such a comparison would serve no useful purpose as the possible differences in the crystalline arrangements of the different specimens due to different mechanical treatment would vitiate any conclusions drawn from the results.

REVIEWS OF BOOKS

Experimental Physics: A Selection of Experiments, by G. F. C. SEARLE. Pp. xiv + 363. (Cambridge University Press.) 16*s*.

In welcoming the publication of another manual of practical physics by the doyen of demonstrators, Dr G. F. C. Searle, we regret to learn that this, the fourth volume in his series of laboratory text-books, is not to be followed by another on experimental electricity and magnetism. From our knowledge of the excellence of the four volumes published we should have known that the completed pentateuch would have been a final authority on laboratory investigations of physical laws and principles. The clarity of exposition and the ample practical details which are given will prevent much wandering in the wilderness on the part of teachers and others who wish to set up any of these original experiments. It is not suggested that such wandering should be condemned as being any more unprofitable in a physical laboratory than, as history shows, it was in the desert of Et Tih, but the great experience and the genius of Dr Searle are so very evident in each one of the experiments, both in the theoretical discussion of a problem and in the design of the apparatus used to investigate it, that these records of his methods of dealing with specific problems stand out as models for all. The present volume contains accounts of experiments in mechanics, elasticity, surface tension, viscosity, heat and sound, none of which occurs in the three manuals already published, together with other experiments which have not been described previously by Dr Searle except in the manuscripts written for use by his students in the Cavendish Laboratory.

A practical example, with numerical data, follows each experiment. It is not always clear that the order of accuracy to which a result is stated is that to be expected from the measurements on which the result is based. For example, in determining the viscosity of water by flow through a capillary tube only two observations of the time of flow of unit mass, t/M, were made. These differ by about two per cent. The mean value is taken and the crude result is worked out. Corrections for calibration of the flow tube, buoyancy, etc., are then made by means of a factor 0·9915, better expressed as (1 − ·0085), and the final result is stated as 0·01066 c.g.s. units. A little more time spent in settling the value of t/M more accurately would have been time well spent. J. H. B.

A Study of Crystal Structure and its Applications, by WHEELER P. DAVEY, Ph.D. Pp. xi + 695, with numerous illustrations. (McGraw Hill Book Co., 1934.) 45*s*. net.

This book is a review of the technique of crystal analysis, and the author states that it is intended for college graduates who wish to read the literature and do independent experimental work. It is based on lectures given over fourteen semesters in the Pennsylvania State College. It is divided into three main sections containing the necessary preliminary information about diffraction and crystal structure, a description of methods of crystal analysis, and a description of its applications in physics, chemistry and metallurgy. The appendices deal with X-ray apparatus, charts for the powder method, and space groups.

A book of this kind may either be a concise review of the main principles of the subject, which refers the student to the original papers for a detailed description of technique, or a comprehensive treatise for the X-ray worker which selects for description the most recent and successful methods and theoretical advances. It seems to the reviewer that the present work does not achieve either of these objects. So long a book falls into the category of the comprehensive treatise, yet in many cases it goes in great detail into descriptions of early methods and ideas without mentioning the simpler and better ones that have since replaced them. This applies both to questions of technique and to theoretical discussions which must surely now be based on quantum mechanics and not on the semi-classical ideas of the Lewis-Langmuir and first Bohr atom, invaluable though these conceptions were in their time. To put it briefly, many sections of the book are years behind the times, and in a subject which is growing so fast this is a serious defect.

The work of Zwicky on secondary structure in crystals has been described at length without reference to the subsequent work of Pauling, Orowan, and Smekal which claims to show that it is based on very erroneous calculations of lattice energy. In the chapter on the powder method the use of molybdenum K rays is described, with bare reference to the vast literature on the use of rays from anticathodes in the calcium-zinc range which has resulted in so enormous an improvement of accuracy. M_0 rays have only been used because X-ray tubes with this anticathode are easily available; they are quite unsuited to accurate work. The description of the structure of Ag_3Al (p. 549) is based on a reference in the *International Critical Tables* of 1926, apparently in ignorance of its solution by Westgren and Bradley two years later. Probably few workers in this field would agree with the author's censure of the new nomenclature of space groups adopted by international agreement, which is widely held to be conspicuously successful in meeting the needs of X-ray analysis. The adoption of its rational system might have saved the author from giving in his list of screw axes (p. 213) only four of the eleven possible types, and listing as monoclinic (p. 21) the face-centred orthorhombic lattice. These are small points, but they are indicative of a want of balance and sense of proportion which mars the book. It contains many suggestive and interesting ideas, and is attractively illustrated. It does not appear to the reviewer, however, to present an accurate picture of the present state of the subject.

Electrolytes, by HANS FALKENHAGEN. Translated by R. P. Bell. Royal 8vo, pp. 346. (Oxford University Press.) 25*s*. net.

The volume gives an admirable account of a subject which is receiving but little attention from physicists in this country. The foundations laid by Faraday have been built upon by Debye and other workers abroad. In the book will be found a comprehensive survey of the modern theory of electrolytes and an indication of the experimental investigations and the data obtained.

The early chapters are devoted to a consideration of the theory of dilute solutions following on the lines laid down by Planck and Arrhenius. This theory is only valid for weak electrolytes, and the author then proceeds to consider strong electrolytes. The behaviour of strong electrolytes shows considerable departures from Arrhenius's theory, and to explain these account must be taken of interionic action. It is only with difficulty that the theory can be extended from the field of dilute solutions to embrace concentrated solutions. Such an extended theory must take into account the polarization, dispersion and repulsive forces between the ions, as well as the forces of interaction between the ions and the solvent molecules. The theoretical treatment of these complicated factors has so far been only partially successful. An account is given of the various attempts that have been made to treat theoretically the properties of concentrated solutions and the question of the true degree of dissociation of an electrolyte is closely connected with this problem. A section of the book is devoted to points connected with the optics of electrolytes and the detection of the undissociated part. It is of interest to note that attempts have been made to study the structure of electrolyte solutions by means of the Raman effect. Prof. R. H.

Fowler contributes an appendix on recent applications of quantum mechanics to the theory of electrode processes.

The volume is to be strongly recommended to students wishing to enter this promising field of enquiry. A word of praise is also due to Mr Bell for the excellence of his translation.

E. G.

Magnetism and Matter, by E. C. STONER, Ph.D. Pp. xv + 575, with 87 diagrams. (Methuen and Co., Ltd., 1934.) 21*s*. net.

Progress in the study of magnetic phenomena has been particularly rapid since the discovery of the electron, but at no period has it been more rapid than in the eight years which have elapsed since the publication of Stoner's *Magnetism and Atomic Structure*. The rate of progress has been such that the new book, *Magnetism and Matter*, is virtually a new work. It takes a very definite and important place in the literature of magnetism, for it occupies a position somewhere between the purely descriptive accounts of experimental facts and procedure and the more rigorous theoretical treatises such as Van Vleck's *Theory of Electric and Magnetic Susceptibilities*.

Consequently one must not turn to Stoner for the most comprehensive experimental details, although complete references and many useful accounts of experimental methods are to be found in his work. The experimental sections have been brought up to date by the inclusion of descriptions of the Sucksmith ring balance (which might, incidentally, be amended in the next edition) and of Kapitza's work with intense magnetic fields. But one must turn to Stoner for an outline of the more important experimental facts and their possible interpretations in the light of modern theories. Thus, those readers who require an easy approach to the present-day theory of paramagnetism will find it here. Here, too, is a readable outline of the modern theory of ferromagnetism and a treatment of the modern aspect of magnetoelastic and magnetoelectric effects. There is very much to be said in favour of such an adequate guide, as a preliminary to the more advanced mathematical treatises and works on modern magnetism.

In his preface the author states that the selection of investigations for detailed consideration has necessarily been somewhat arbitrary, but there must be singularly few competent persons who will find fault with his selection. The book is an important addition to the standard works on physics and is heartily commended to the Fellows of the Physical Society.

L. F. B.

The Kinetic Theory of Gases (Second Edition), by L. B. LOEB. Pp. xx + 687. (McGraw-Hill Book Co., 1934.) 36*s*. net.

In reviewing a *Kinetic Theory* of the scope of this book some comparison with the well-known treatise of Jeans seems almost unavoidable, and in any case there is no reason why so useful a comparison should be avoided.

Though much of the same ground is covered in the two treatises, Loeb lays much more emphasis than Jeans on the experimental side of the kinetic theory, and he deals exhaustively with a great many modern applications and developments of the subject. On the mathematical side the treatment is less systematic and less elegant than that of Jeans. To use a possibly far-fetched analogy, Jeans climbs his mathematical hills quietly and easily in top gear and then in special physical chapters tells his humbler readers exactly what is visible from the hill-tops. Loeb, on the other hand, takes the experimental physicist and chemist with him up such hills as he elects to climb. The load being heavier, this choice makes for a certain amount of gear-grinding, but on the whole there is much to be said for it.

Prof. Loeb clearly writes with a real enthusiasm for his subject. The book is of special value in that it combines in a single volume accounts of the classical kinetic theory and of the most recent developments. It thus provides, both for undergraduate and for advanced students, a valuable compendium of information, much of which is not elsewhere available in text-book form. In particular mention must be made of the chapter, nearly 100 pages long, on the applications of kinetic theory to problems in electrical conduction through gases, where Prof. Loeb can write freely as a leading authority.

There are, unfortunately, some minor blemishes in the work. There are some slips which should not have appeared in a second edition, and some errors—for example, the derivation of Poiseuille's formula on the assumption that a gas behaves as an incompressible fluid. There is also some carelessness in the writing of proper names. Most of these are too trivial for comment, but it is not in human nature to overlook the reference in the index to "Sir W. Rayleigh"—which suggests an ingenious telescoping, across almost exactly three centuries, of a great physicist and a great Elizabethan seaman. It must, however, be added that if Prof. Loeb is slightly inaccurate in some of his references, he is precise and very generous in his acknowledgments.

Prof. Loeb deserves credit for emphasizing the historical side of the subject, but the weight of evidence is against some of his historical notes. Brown, for instance, would have been much in advance of the other botanists of his time if in 1827 he had been led (p. 390) "to venture the idea that the (Brownian) motion was due to the unequal bombardment of the particles on various sides by the molecules of the liquid executing their heat motions and that they were therefore a manifestation of molecular heat motions." The kinetic-molecular explanation is generally attributed to Wiener (1863), and in his paper of 1829 Brown speaks of "motions for which I am unable to account." There is little evidence of any strong interest in the motions, and still less of any "violent controversy," for some years after Brown's discovery.

Again (pp. 84, 139) there has been some misapprehension concerning the date of publication of Waterston's paper on "The Physics of Media that are Composed of Free and Perfectly Elastic Molecules in a State of Motion." This was communicated to the Royal Society in 1845 by Beaufort, but it was not until nearly fifty years later that Rayleigh, then Secretary, rescued it from the Archives and caused it to be printed in the *Phil. Trans.* for 1892. It is not likely that Rayleigh, who was born in 1842, had in 1845 any very pertinent comment to make on the importance of the paper. If he had, his precocity would have been almost as remarkable as the vitality and versatility of Robert Brown, who was in his fifties when he discovered the Brownian movements, but who is shown in the index as O. W. Richardson's collaborator more than eighty years later.

These minor sources of irritation or amusement apart, this is a very good book.

H. R. R.

Bessel Functions for Beginners, by N. W. McLachlan, D.Sc., M.I.E.E. Pp. xi + 192. (Clarendon Press, 1934.) 15*s*. net.

The object of this book is to provide engineers and engineering students with a course of Bessel functions sufficiently advanced to enable them to employ the functions in practice in their physical and engineering researches. The book is well written and a great deal of labour has obviously been expended on it. Besides the worked-out examples there are 600 other examples set with answers. It will thus be as useful to students of pure as to students of applied mathematics. A graduate student looking out for a mathematical or physical subject as a thesis for a doctor's degree would find this a suggestive book. We can thoroughly recommend it.

A. R.

Les Appareils à Fil Chaud, by E. G. RICHARDSON. Pp. 67. (Paris: Gauthier-Villars, 1934.) 20 fr.

This memoir contains in an expanded form the substance of a lecture given during 1932 by its author at the Institut de Mécanique des Fluides of the University of Paris. It is divided into two parts, the first of which contains an outline of the history and theory of hot-wire anemometers together with instructions for their use including particulars of precautions and corrections to be observed and made. In the second part the results of investigations making use of such instruments are described: these investigations are concerned with such phenomena as the flow around aerofoils, vortex formation and turbulence, the distribution of air-velocity in acoustical resonators, flow in tubes, and the viscosity of colloids.

The book summarizes conveniently the advances made by the author (who in many of his researches has collaborated with Dr Piercy of the Aeronautical Department at East London College and Dr Tyler of Leicester Technical College), while a bibliography of some sixty references ensures that adequate weight is given to the results of other workers in this field. All those who contemplate using a hot wire will be well advised to economize in both time and trouble by securing a copy of this book. E. J. I.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

Vol. 47, Part 3 May 1, 1935 No. 260

621.396.674

ON A THEORY OF THE ACTION OF RECTANGULAR SHORT-WAVE FRAME AERIALS

ERRATA

"A new precision colorimeter", by John J. Manley, M.A., D.Sc. Oxon., *Proc. phys. Soc.* **47**, 70 (1935), line 1:

for 0·1 mm. *read* 1·0 mm.

"The propagation of medium radio waves in the ionosphere", by D. F. Martyn, Ph.D., A.R.C.Sc., F.Inst.P., *Proc. phys. Soc.* **47**, 325–334 (1935):

for ¶ *read* π

of the frames are used.

§ 1. THE ACTION OF RECTANGULAR SHORT-WAVE FRAMES

A current is generally produced in a wireless frame aerial either because the frame is fed by a local high-frequency oscillator or because passing electromagnetic waves are incident upon it. Either of these primary sources of energy will cause an initial current to be set up in the frame. If the frame be rectangular and if its dimensions be comparable with the wave-length of the initial current, then two other supplementary or component currents (which are otherwise negligible) will contribute appreciably to the final or resultant current produced in the frame.

One of these supplementary currents is due to the fact that the initial current in one limb of the frame at any point *P*, say, contributes to the resultant current in the opposite limb of the frame at *P*′, say, by direct radiation across the frame, whilst the

THE PROCEEDINGS OF
THE PHYSICAL SOCIETY

VOL. 47, PART 3 May 1, 1935 No. 260

621.396.674

ON A THEORY OF THE ACTION OF RECTANGULAR SHORT-WAVE FRAME AERIALS

BY PROFESSOR L. S. PALMER, D.Sc., Ph.D., M.I.E.E., F.Inst.P.
AND
DENIS TAYLOR, M.Sc.
The University College, Hull

Received September 14, 1934. *Read February* 15, 1935.

ABSTRACT. In a previous approximate theory of the action of rectangular frame aerials used with waves comparable in length to the frame-dimensions, it was assumed that the frame current would be a maximum when the initial current produced by a local oscillator or a passing electromagnetic wave was in phase with a supplementary current produced by the mutual reactions between the currents in opposite limbs of the frame. The co-phasing of these component currents was accomplished by adjusting the dimensions of the frame, a process for which the term "formatizing" was suggested.

The present paper develops this theory, and the critical dimensions for rectangular transmitting and receiving frames are exactly determined. It is found that the maximum error arising from the use of the approximate formulae previously given does not exceed 4·7 per cent and is usually less than 3 per cent. These errors are not greater than the experimental errors which arise when short waves comparable in length to the dimensions of the frames are used.

§ 1. THE ACTION OF RECTANGULAR SHORT-WAVE FRAMES

A CURRENT is generally produced in a wireless frame aerial either because the frame is fed by a local high-frequency oscillator or because passing electromagnetic waves are incident upon it. Either of these primary sources of energy will cause an initial current to be set up in the frame. If the frame be rectangular and if its dimensions be comparable with the wave-length of the initial current, then two other supplementary or component currents (which are otherwise negligible) will contribute appreciably to the final or resultant current produced in the frame.

One of these supplementary currents is due to the fact that the initial current in one limb of the frame at any point P, say, contributes to the resultant current in the opposite limb of the frame at P', say, by direct radiation across the frame, whilst the initial current at P also contributes to the resultant current at P by reflection of its radiation back to P from the opposite limb of the frame.

The other supplementary current arises from the fact that the initial disturbance at P causes a wave to travel round the frame-perimeter back to P where it, also, will contribute to the initial disturbance.

In the previous communications (1, 2, 3) an approximate theory of the first or radiation effect has been given, whilst the second effect has been discussed in a paper on the current-distribution round a short-wave frame aerial (4). It is the object of the present note to consider the magnitude of the errors arising from the approximate theory of the radiation effect which has been described in the communications (1, 2, 3) referred to above.

§ 2. THEORETICAL CONSIDERATIONS: TRANSMISSION

I, I_0, ω

i, i_0, ρ_t

i'

i_0', ρ_t'

λ, D

Consider first the case of a frame actuated by a local oscillator, and let the initial current per centimetre at P in one limb of the frame be I, where $I = I_0 \sin \omega t$. This initial current will not be that actually measured because it will be more or less reinforced by the supplementary currents produced by the radiation effects referred to above. In the case of a transmitting frame, let this initial current I at P produce a supplementary current i at P given by $i = i_0 \sin(\omega t + \rho_t)$ per centimetre and also a supplementary current i' at P', in the opposite limb of the frame, given by $i' = i_0' \sin(\omega t + \rho_t')$ per centimetre. The amplitudes i_0 and i_0' are dependent on I_0, on the wave-length λ in air, and on the distance D between the opposite limbs of the frame. ρ_t and ρ_t' are phase-changes between i and i' respectively and the initial current I. These phases are dependent on both time and space. We may write the supplementary currents in the general form

$$f(\lambda D)\, I_0 \sin\left[\omega t + 2\pi \int (\nu\, dt - ds/\lambda)\right],$$

ν, t, s

I_R

where ν is the wave frequency and dt and ds are increments of time and space respectively. The initial and supplementary currents at P will have a resultant current I_R, the amplitude of which will depend on the amplitudes of and the phase-difference between the component currents. If i_0 and i_0' were constant and independent of D, then, for a given wave-length, a maximum resultant current would occur when the components I and i at P were in phase and when the components I at P and i' at P' were π out of phase. The approximate theory already given assumed that for a given wave-length the variation of i_0 or i_0' with D was negligible for values of D not very different from that particular value for which the two components were in or out of phase. Thus it was concluded* that the frame current would be a maximum when

$$\rho_t = \pm 2n\pi \qquad \text{......(1a)}$$

and

$$\rho_t' = \pm (2n' + 1)\,\pi \qquad \text{......(1b)},$$

n, n'

where n and n' are integers.

For the particular case of a transmitting frame the two components I and i recombine at P after a time interval $2t$, say, during which the initial component I

---

* See references (2) and (1) respectively.

at P has continued, whilst the supplementary component i at P has resulted from radiation which has travelled across the frame and back through a total distance of $2D$. Thus at point P and time $2t$ the initial component I has suffered a phase-change of $2\pi . 2\nu t$ whilst the supplementary component i has also suffered a phase-change of $2\pi . 2\nu t$ together with other phase-changes due to the distance $2D$ and to the mechanism of emission and absorption of radiation by the limbs of the frame. Thus the resultant phase-change may be considered as that due to the time interval $2t$ together with that due to the "effective" distance traversed by the radiation.

It has been shown$^{(1,2,3)}$ that this "effective" distance is given by

$$2\left(D - \frac{\lambda}{2\pi}\tan^{-1}\frac{a^2-1}{a}\right),$$

where $a = 2\pi D/\lambda$. Hence, for the case of a transmitting frame, the phase-difference a
between the supplementary current i at P and the initial current I at P is given by

$$\begin{aligned}\rho_t &= 2\pi\left\{2\left[\nu t - \frac{1}{\lambda}\left(D - \frac{\lambda}{2\pi}\tan^{-1}\frac{a^2-1}{a}\right)\right] - 2\nu t\right\}\\ &= -2(a-\phi),\end{aligned}$$

where $\phi = \tan^{-1}\dfrac{a^2-1}{a}$. If the assumption that $\rho_t = \pm 2n\pi$ be correct, then we have ϕ

$$-2(a-\phi) = \pm 2n\pi$$

or

$$\tan(a \pm n\pi) = (a^2-1)/a \qquad \text{......(2}a\text{)}.$$

Similarly the two components at P' will recombine after a time interval t during which the supplementary component i' will have resulted from radiation which has travelled across the frame a distance D. Thus the phase-difference between the supplementary current i' at P' and the initial current I at P is ρ_t', where

$$\begin{aligned}\rho_t' &= 2\pi\left\{\left[\nu t - \frac{1}{\lambda}\left(D - \frac{\lambda}{2\pi}\tan^{-1}\frac{a^2-1}{a}\right)\right] - \nu t\right\}\\ &= -(a-\phi).\end{aligned}$$

Hence, if $\rho_t' = \pm(2n'+1)\pi$, we have

$$\tan[a \pm (2n'+1)\pi] = (a^2-1)/a \qquad \text{......(2}b\text{)}.$$

Since $\rho_t = 2\rho_t'$ these equations—and equations (1)—can only be satisfied simultaneously if n be an odd integer. Consequently, only alternate solutions of equation (2a) give what were, for convenience of reference, called the first formatizing conditions for transmission,* that is, the conditions giving the critical dimensions of the frame for maximum current from a local oscillator of constant voltage output. By introducing the arbitrary parameter ψ the final formatizing conditions for rect- ψ
angular transmitting frames were given as

$$\left.\begin{aligned}&\tan(a-\psi) = (a^2-1)/a\\ \text{and}\quad &\tan(a'+\psi) = (a'^2-1)/a'\end{aligned}\right\} \qquad \text{......(3)},$$

where $a = 2\pi W/\lambda$ and $a' = 2\pi H/\lambda$, W and H being the frame-width and height respectively.

---

* See reference (2), p. 66.

It is now desirable to examine the function $f(\lambda D)$ in order to see how i_0 and i_0' vary with D. This will enable us to determine to what extent the assumptions underlying equations (1) and (2) are justified and will also enable the magnitude of the errors involved in equations (3) to be calculated.

If the initial current at P in one limb of the rectangular frame be $I_0 \sin \omega t$ per cm., then the supplementary current i' per cm. produced in the parallel opposite limb a distance D away is given by the equation

$$i' = \frac{J \, . \, I_0}{D^3} \sqrt{[(a^2-1)^2+a^2]} \sin[\omega t-(a-\phi)],$$

J where J is a constant.* For convenience of calculation this expression may be written in the form

$$\begin{aligned} i' &= KI_0 \sqrt{(1/a^2-1/a^4+1/a^6)} \sin[\omega t-(a-\phi)] \\ &= i_0' \sin(\omega t+\rho_t'). \end{aligned}$$

K R Here K is another constant for a given wave-length. With i' in amperes per centimetre and D and λ in centimetres, the constant $K=120\pi^2/\lambda^2 R$, where R is the resistance in ohms per centimetre of the frame wire. The numerical value of K for the short waves under consideration and for suitable copper wire is of the order of unity.† It is shown below that the exact value of K is of little importance in connection with the present problem.

The field radiated by the current i' is responsible for producing the supplementary current i at P in the original wire. The relation between i' and i will be similar to that between I and i', and consequently

$$\begin{aligned} i &= Ki_0' \sqrt{(1/a^2-1/a^4+1/a^6)} \sin[\omega t-2(a-\phi)] \\ &= K^2(1/a^2-1/a^4+1/a^6) I_0 \sin[\omega t-2(a-\phi)] \\ &= i_0 \sin(\omega t+\rho_t) \qquad \text{......(4).} \end{aligned}$$

For convenience in determining the magnitude of the errors involved in the use of equations (1) or (2), the discussion will be restricted to a consideration of the supplementary current i at P, although the same results can be obtained by considering the supplementary current i' at P'.

In figure 1, $i_0/K^2 I_0$ is plotted against a (i.e. $2\pi D/\lambda$), whilst ρ_t is plotted against a in the full-line graph of figure 2. By plotting corresponding data from these two graphs, the spiral vector diagram of figure 3 is obtained. Figure 3*b* is figure 3*a* enlarged. Any point on this spiral, when multiplied by the scalar constant $K^2 I_0$, represents vectorially the supplementary current i for any particular value of a. In the theory already given any small element of this spiral was assumed to be part of a circle for values of a greater than about 1·25, i.e. for values of D greater than 0·2λ. If this were true, it is obvious that the resultant current I_R would be greatest when its components I and i were exactly in phase, and equations (2*a*) and (3) would be justified. Because i_0 is a function of D (for a constant wave-length) these equations are only approximately true, and the exact value of the amplitude of I_R, the resultant current, is given by

---

* See reference (1), p. 195, where the nomenclature is slightly different from that used in the present paper.

† According to some recent measurements by Walmsley (*J. Instn elect. Engrs*, **69**, 311) the resistance per cm. of 1/0·044 copper wire for a wave-length of about 15 metres is 0·0055 Ω. approximately. Therefore for this wire $K=0·6$.

$$ {}_0I_R = \sqrt{[I_0^2 + i_0^2 + 2I_0 i_0 \cos 2\,(a-\phi)]} \qquad \text{......(5).} $$

${}_0I_R$

On substituting in equation (5) the value of i_0 given in equation (4), we get

$$ {}_0I_R = I_0 \sqrt{[1 + K^4 (1/a^2 - 1/a^4 + 1/a^6)^2 + 2K^2 (1/a^2 - 1/a^4 + 1/a^6) \cos 2\,(a-\phi)]}. $$

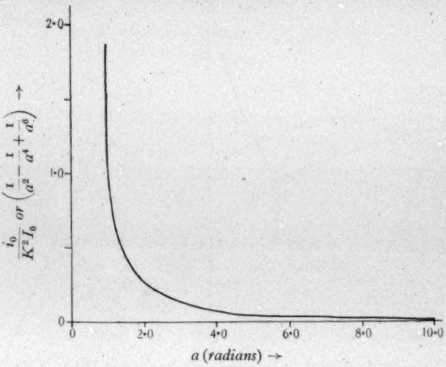


Figure 1. Graph showing the variation in amplitude of the first supplemental current.

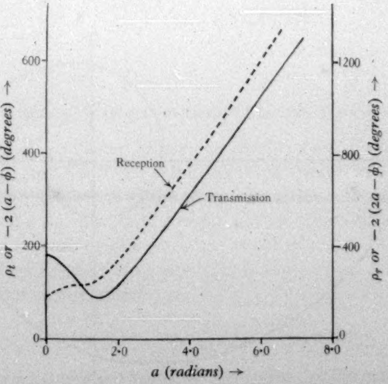


Figure 2. Graph showing the variation in phase of the first supplemental current.

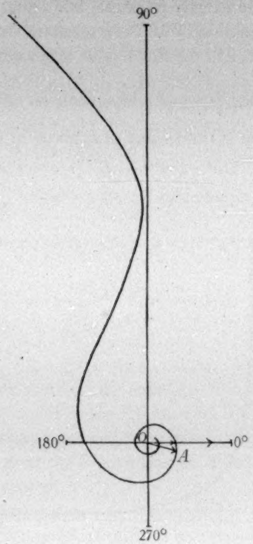


Figure 3*a*. Vector diagram showing the variation in amplitude and phase of the first supplemental current in transmission.

$$r=\left(\frac{1}{a^2}-\frac{1}{a^4}+\frac{1}{a^6}\right),\quad \theta=-2\,(a-\phi).$$

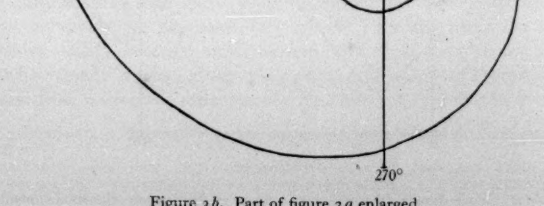


Figure 3*b*. Part of figure 3*a* enlarged.

On equating $d\,({}_0I_R)/d\,(a)$ to zero we get the conditions for maximum or minimum resultant current at the point P in the frame, namely

$$-2\,(a-\phi)=\left\{\tan^{-1}\left[\frac{1}{a}-\frac{2}{a\,(a^2-1)^2}\right]+\sin^{-1}\frac{K^2\,[(a^2-1)^2+2]\,[(a^2-1)^2+a^2]}{a^6\sqrt{[(a^2-1)^2+2]^2+[a\,(a^2-1)^2]^2}}\right\}\pm 2n\pi$$
$$=\tan^{-1}A+\sin^{-1}B\pm 2n\pi \qquad \text{......(6).}$$

This may be compared with equation (2a) which is

$$-2\,(a-\phi)=\pm 2n\pi.$$

In order to solve these transcendental equations it is convenient to write them respectively in the form

$$a\pm n\pi=\phi-\tfrac{1}{2}\,(\tan^{-1}A+\sin^{-1}B)$$

and

$$a\pm n\pi=\phi.$$

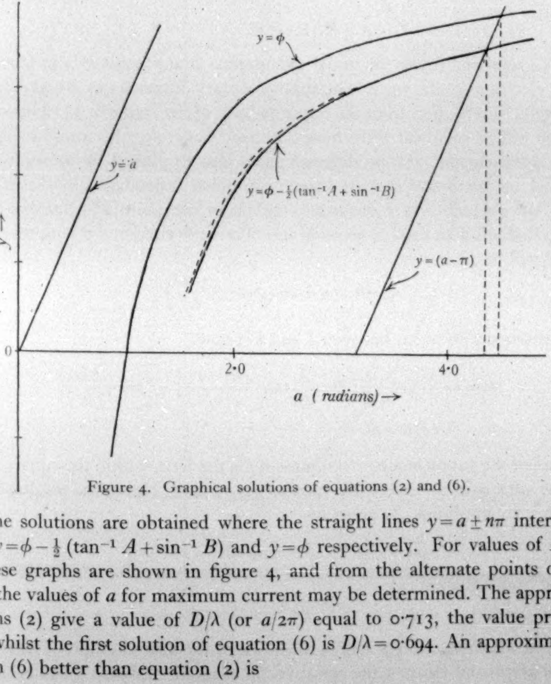


Figure 4. Graphical solutions of equations (2) and (6).

Then the solutions are obtained where the straight lines $y=a\pm n\pi$ intersect the curves $y=\phi-\frac{1}{2}\,(\tan^{-1}A+\sin^{-1}B)$ and $y=\phi$ respectively. For values of K equal to 1 these graphs are shown in figure 4, and from the alternate points of intersection the values of a for maximum current may be determined. The approximate equations (2) give a value of D/λ (or $a/2\pi$) equal to 0·713, the value previously given; whilst the first solution of equation (6) is $D/\lambda=0{\cdot}694$. An approximation to equation (6) better than equation (2) is

$$-2\,(a-\phi)\doteqdot\left(\tan^{-1}\frac{1}{a}+\sin^{-1}\frac{K^2}{a^3}\right)\pm 2n\pi \qquad \text{......(7).}$$

The graph of $y=\phi-\frac{1}{2}\,(\tan^{-1}a^{-1}+\sin^{-1}a^{-3})$ is shown by the dotted curve in figure 4, and the resulting critical values of D/λ are hardly distinguishable from those given by equation (6). The very small value of the arc sin term shows that the correcting factor is practically independent of the precise value of K as long as K does not differ greatly from unity.*

On substituting 0·694 for D/λ in equation (6) we get the value of the correcting term $\frac{1}{2}(\tan^{-1} A + \sin^{-1} B)$ which is found to be equal to 12·8°. Thus the decrease in I_R due to this phase angle between the components I and i is more than compensated by the increased value of the amplitude of i (i.e. i_0); whilst, for phase angles greater than 12·8°, the further increase in i_0 is insufficient to compensate for the reduction in I_R arising from the larger phase-difference between its component currents. This critical value of i_0 (divided by the scalar constant $K^2 I_0$) is shown in figure 3 by the vector OA.

§ 3. THEORETICAL CONSIDERATIONS: RECEPTION

With a receiving frame the initial disturbance at any point P is produced by an incident electromagnetic wave. The supplementary currents i at P and i' at P' due, respectively, to reflection from the opposite limb of the frame or to radiation across the frame will be identical with those discussed above for the transmitting frame; but the initial current will be different from that produced by a local oscillator. The initial disturbance at P, with which i combines, is produced by that part of the incident wave which was a distance $-2D$ from the point P when the reflected radiation first left P at time $2t$ seconds ago. Thus the current I due to this retarded potential will be given by

$$I_0 \sin\left[\omega t + 2\pi\left(2\nu t - \frac{-2D}{\lambda}\right)\right]$$

ρ_r and the phase-difference ρ_r between I and i is now

$$\rho_r = 2\pi\left\{\left[2\nu t - \frac{2}{\lambda}\left(D - \frac{\lambda}{2\pi}\tan^{-1}\frac{a^2-1}{a}\right)\right] - \left(2\nu t + \frac{2D}{\lambda}\right)\right\}$$
$$= -2\,(2a-\phi).$$

The phase of the supplementary component i is the term within the square brackets and is the same as in the case of the transmitting frame. Hence the resultant current I_R at P has an amplitude ${}_0I_R$ given by

$${}_0I_R = \sqrt{[I_0^2 + i_0^2 + 2I_0 i_0 \cos 2\,(2a-\phi)]} \qquad \ldots\ldots(8)$$

in which, from equation (4),

$$i_0 = K^2 I_0\,(a^{-2} - a^{-4} + a^{-6}).$$

The way in which ρ_r changes with a is shown in the dotted graph of figure 2, and from this graph and figure 1 the spiral vector diagram of figure 5 has been drawn.

On substituting in equation (8) for i_0 as before, and equating the differential with respect to a to zero, we get equation (9), the odd solutions of which will give the values of D/λ for maximum receiving frame current:

---

* See footnote † on p. 380.

$$-2(2a-\phi)=\left\{\tan^{-1}\left[\frac{1}{2a}-\frac{a^2-4}{4a(a^2-1)^2+2a^3}\right]\right.$$
$$\left.+\sin^{-1}\frac{K^2[(a^2-1)^2+2][(a^2-1)^2+a^2]}{a^6\sqrt{[\{(a^2-1)^2+2\}^2+\{2a(a^2-1)^2+a^3\}^2]}}\right\}\pm 2n\pi \quad \ldots\ldots(9)$$
$$=\tan^{-1}A'+\sin^{-1}B'\pm 2n\pi.$$

A', B'

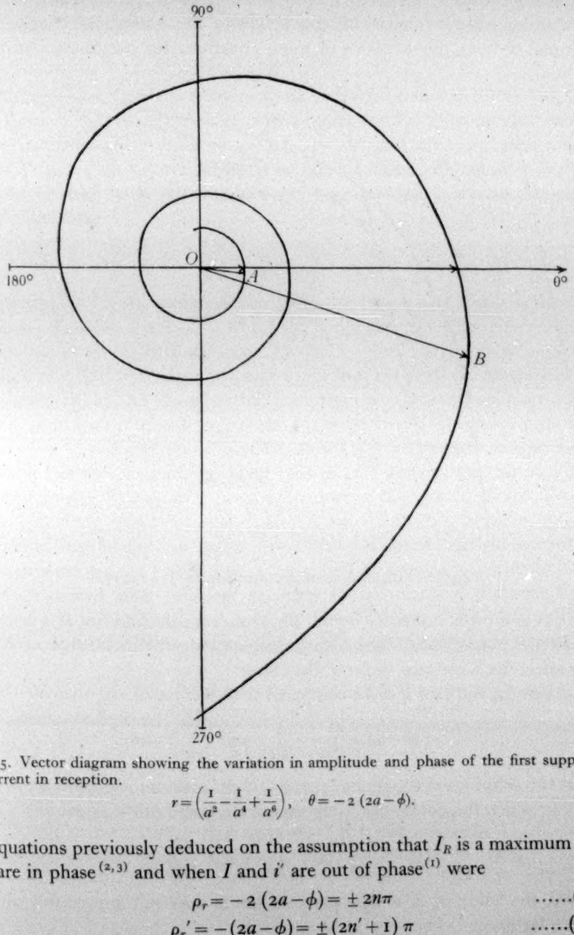


Figure 5. Vector diagram showing the variation in amplitude and phase of the first supplemental current in reception.

$$r=\left(\frac{1}{a^2}-\frac{1}{a^4}+\frac{1}{a^6}\right),\quad \theta=-2(2a-\phi).$$

The equations previously deduced on the assumption that I_R is a maximum when I and i are in phase ⁽²˒³⁾ and when I and i' are out of phase ⁽¹⁾ were

$$\rho_r=-2(2a-\phi)=\pm 2n\pi \quad \ldots\ldots(10a)$$

and

$$\rho_r'=-(2a-\phi)=\pm(2n'+1)\pi \quad \ldots\ldots(10b).$$

By introducing an arbitrary phase angle ψ and also the effective frame-width
W, H, γ $W \cos \gamma$ and the effective frame-height $H \sin \gamma$ for waves incident at an angle γ to the frame, the formatizing conditions for rectangular receiving frames were given by

$$\left.\begin{aligned} &\tan\left[a\left(1+\cos\gamma\right)-\psi\right]=\left(a^2-1\right)/a \\ \text{and}\qquad &\tan\left[a'\left(1+\sin\gamma\right)+\psi\right]=\left(a'^2-1\right)/a' \end{aligned}\right\} \qquad \ldots\ldots(11),$$

where a and a' refer to the width and height of the frame respectively. With γ and ψ equal to zero, the solutions of these equations for maximum currents are $W/\lambda=0{\cdot}331$ and $0{\cdot}860$ and $H/\lambda=0{\cdot}713$. This result for the value of H/λ is the same as that for the transmitting frame because, when $\gamma=0$, the incident wave does not directly affect the horizontal limbs of the frame.

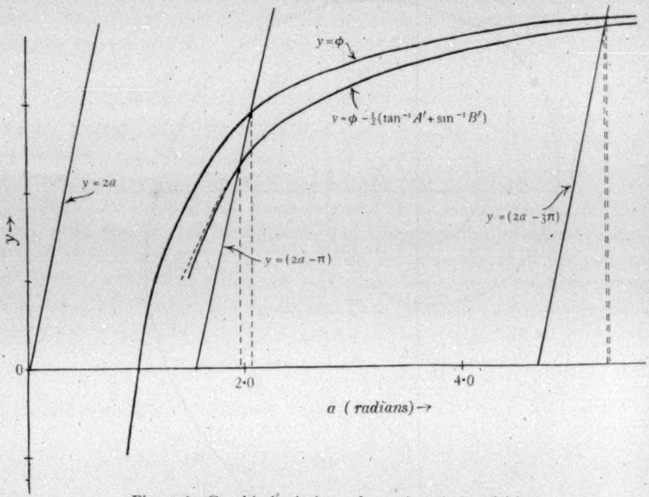


Figure 6. Graphical solutions of equations (10) and (9).

Equations (9) and (10) may be compared with the useful approximate equation

$$-2\left(2a-\phi\right) \doteqdot \left(\tan^{-1}\frac{1}{2a}+\sin^{-1}\frac{K^2}{2a^3}\right) \pm 2n\pi \qquad \ldots\ldots(12).$$

The first two solutions of equations (9) and (10) for maximum current are shown in figure 6, in which the dotted curve represents the graph of the equation

$$y=\phi-\frac{1}{2}\left(\tan^{-1}\frac{1}{2a}+\sin^{-1}\frac{1}{2a^3}\right).$$

As before, the value of K within practical limits does not appreciably alter the correction factor.

The straight lines are the graphs of $y=(2a \pm n\pi)$. The solutions previously obtained from equation (10) or equation (11), namely $D/\lambda=0{\cdot}331$ and 0·860, may now be compared with the new results from equation (9). These are $D/\lambda=0{\cdot}316$ and 0·856 when $n=1$ and 3 respectively.

On substituting these new values in equation (9), the correcting terms on the right-hand side are found to be 18·0° and 8·4° respectively. Thus the critical values of i which, together with I, will give a maximum resultant current are

$$i_0 \sin(\omega t - 342^\circ) \quad \text{and} \quad i_0 \sin(\omega t - 1071{\cdot}6^\circ)$$

respectively. These supplementary currents are proportional to the vectors OB and OA in figure 5. Alternate points of intersection must be taken because, when the component currents at P differ by 4π, 8π, 12π, etc. and satisfy equation (10*a*) they differ at P' by 2π, 4π, 6π, etc. which is not in agreement with equation (10*b*). This point was referred to in the discussion of equation (2*a*) in connection with the transmitting frame.

§4. CONCLUSION

Experimental measurements of the critical or formatizing dimensions of transmitting and receiving frames (for which the frame currents attain maximum values for given input voltage) yield values of D/λ which agree closely with those calculated from the theoretical formulae discussed above, but the errors of measurements are usually greater than the errors arising from the use of the approximate equations (2) and (10) instead of equations (6) and (9) respectively. The percentage error in the first value of D/λ given by equations (2) or (3) is 2·8 per cent, whilst the errors of the first two solutions of equations (10) or (11) are 4·75 per cent and 0·47 per cent respectively. These errors are independent of the value of K for all practical frames.

It would therefore appear that the best shapes of rectangular short-wave frames for transmission and reception (that is, the critical dimensions for formatized frames) may be calculated with sufficient accuracy for all practical purposes from the theoretical equations (2) and (10) based on the co-phasing of the initial and supplementary currents, whilst for more accurate calculations equations (7) and (12) are useful.

REFERENCES

(1) Palmer. *Proc. roy. Soc.* A, **136**, 193 (1932).
(2) Palmer and Taylor. *Proc. phys. Soc.* **46**, 62 (1934).
(3) —— —— *Proc. Inst. Radio Engrs*, N.Y., **22**, 93 (1934).
(4) Palmer, Taylor and Witty. *Proc. phys. Soc.* **46**, 76 (1934).

621.396.674

THE CURRENT VARIATIONS IN A SHORT-WAVE SQUARE FRAME AERIAL REVOLVING IN ITS OWN PLANE

By Professor L. S. PALMER, D.Sc., Ph.D., M.I.E.E., F.Inst.P.

AND

ROY WITTY, B.Sc., A.Inst.P.

The University College, Hull

Received November 14, 1934. *Read February* 15, 1935.

ABSTRACT. From a theory of the action of short-wave frame aerials outlined in previous communications it was concluded that the current at any point in a rectangular frame aerial should vary with time as the frame revolved in its own plane. If the current-measuring instrument moves round with the frame, complications arise because the instrument records not only the current-variations due to changes in what was called the degree of formatization of the frame, but also those due to the fact that the instrument passes in succession through the current nodes and antinodes which have been shown to be fixed in space and not to revolve with the frame. Hence current-measurements have been made, firstly with the frame fixed and the instrument moving round the frame, and secondly with the instrument fixed in space as the frame revolved. The apparatus and methods employed are briefly described. The current-variations observed in the two sets of experiments conform approximately to those predicted from the theory. The results lead in general to the conclusion that when a square frame aerial is used with short waves and revolves in its own plane then, in addition to spatial current-variations round the perimeter of the frame due to the formation of fixed current nodes and antinodes, there may also be temporal current variations at any fixed point due to the fact that the frame may become alternately formatized and deformatized as it revolves.

§ 1. INTRODUCTION

According to the familiar view that the e.m.f. in a receiving frame aerial is proportional to the rate of change of the magnetic flux interlinked with it, the current should vary as the frame rotates about an axis in its plane, but there should be no current variations as a receiving frame revolves about an axis perpendicular to its plane, that is, as the frame revolves in its own plane.

It has, however, been shown(1) that current variations are to be expected in the latter case if the frame be rectangular and if its dimensions be comparable with the length of the wave. This arises because other factors are involved which modify the simple theory indicated above. In particular, at any point in the frame the e.m.f. not only is produced by the changing magnetic flux of the incident wave, but also is influenced firstly by radiation from adjacent parts of the frame and secondly by the circulation of current round the frame perimeter. It was shown that the first of these supplemental effects (the radiation effect) and consequently the magnitude of the resulting frame current depends on the ratio of the frame-dimensions to the wave-length and on the angle of incidence of the wave upon the frame sides, the direction of propagation being in the plane of the frame—that is, on the shape and orientation of the frame; whilst the second supplemental, or current-distribution, effect depends only on the ratio of the frame-perimeter to the wave-length and is independent of the shape and orientation of the frame.

A previous communication(2) dealt with this latter or current-distribution effect when not complicated by the former or radiation effect. The isolation of the two effects was accomplished by utilizing frames so shaped that they were deformatized* for all angles of incidence with the particular wave-lengths in use. Furthermore it was shown that the positions of the current nodes and antinodes were determined by two currents circulating round the frame in opposite directions and that the resulting nodes and antinodes were fixed in space with reference to the position of the transmitter.

The present paper deals with the radiation effect isolated from the current-distribution effect. It was thought that a study of the radiation effect by itself should confirm the existence of those current-variations which have been anticipated(1) as a result of revolving a short-wave rectangular frame aerial in its own plane.

§ 2. THEORETICAL CONSIDERATIONS

When a rectangular frame is formatized it is so shaped that the current at any point in the frame becomes abnormally large compared with what it would be for a frame of the same area but of arbitrary shape. The necessary conditions have been termed, for convenience, the "first formatizing conditions"(3)†. From them it follows that a square frame, if formatized when the wave-front is parallel to one side (when it is said to be in the "square position"), will become deformatized when the frame revolves through 45° into the "diamond position", the current at any point becoming less; and the reverse will occur if the frame be formatized for reception in the diamond position.

First, we may consider what current-variations are to be expected if the measuring instrument moves round a fixed frame. The instrument will pass in succession through the current nodes and antinodes spaced round the perimeter and so will record the current-distribution already discussed elsewhere(2). If, however, the instrument be fixed in the frame and both revolve together, then the observed current-variations will differ from those recorded with a fixed frame. This arises because the instrument not only will pass in succession through the spacially fixed current nodes and antinodes as before, but also will record current-variations

---

[*] That is, the ratios of the frame-dimensions to the wave-length were such that, at any given point of the frame, the supplemental e.m.f. due to radiation from other parts of the frame did not augment or reinforce the initial e.m.f. due to a local oscillator or an incident wave. See reference (2), pp. 198, 199.

[†] See equations (1) below.

due to the fact that the frame may become alternately formatized and deformatized as it revolves. For certain wave-lengths, current maxima due to this cause will occur whenever the frame is in the square position, whilst for other wave-lengths maxima will occur when it is in the diamond position. Thus the current measured at any point will depend on the combination of the two different effects discussed above.

The isolation of the radiation effect may be accomplished by measuring the current at any fixed point in space irrespective of the position of the point on the frame. The current at this point will vary with time as the revolving frame becomes alternately formatized and deformatized, but it will not vary as a result of the non-uniform current-distribution round the frame. This follows from the fact that this distribution is fixed in space and is dependent only on the position of the transmitter. Consequently, if it can be so arranged that the current is always measured at the same point in space with respect to the transmitter, then any current-variation which may be recorded will be due to the fulfilment or non-fulfilment of the first formatizing conditions as the frame revolves.

Thus, this method of using a current-measuring instrument fixed in space enables the temporal current changes due to the first formatizing conditions to be studied without any disturbing effect arising from the non-uniform spacial distribution of the current round the frame.

It has already been shown* that the formatizing conditions arising from the mutual radiation between parallel limbs of a rectangular frame are given approximately† by the odd solutions of the equations:

$$\left.\begin{aligned} \tan\,[a\,(1+\cos\gamma)-\psi] &= (a^2-1)/a \\ \tan\,[a'\,(1+\sin\gamma)+\psi] &= (a'^2-1)/a' \end{aligned}\right\} \qquad \ldots\ldots(1),$$

a, W, a', H, λ
γ
ψ

where $a = 2\pi W/\lambda$, $a' = 2\pi H/\lambda$, W/λ and H/λ are the ratios of the frame-width and the frame-height respectively to the wave-length, γ the angle of incidence of the wave-front on the frame and ψ an arbitrary phase angle.

The alternative solutions of these equations give the critical frame-dimensions for maximum and minimum current. The values of a and a' were found for various values of ψ by a graphical method and curves were given showing the relation between W/λ and H/λ for different angles of incidence γ of the wave on the frame. The first few curves for both maximum and minimum current conditions for $\gamma = 0^\circ$ and 45° are given in figures 1 (*a*) and 1 (*b*) respectively. The critical wave-lengths for a square frame, with which this paper is concerned, can be obtained from these curves by drawing a line through the origin at 45° to the axes and noting the values of W/λ (or H/λ) for the points where the straight line cuts the curves. In this way the wave-lengths for both maximum and minimum current conditions can be deduced for a square frame in the square position ($\gamma = 0^\circ$) or in the diamond position ($\gamma = 45^\circ$).

We may now consider what current-variations are to be expected when a square frame revolves in its own plane, the current being measured by an instrument which

---

[*] Reference (1), p. 197.

[†] See preceding paper, p. 377 of this volume.

is always inserted in the frame at that point which is (say) nearest to the transmitter. For a wave-length for which the frame is always deformatized when at any angle to the wave-front, the current measured should show no temporal variation as the frame revolves. This conclusion has already been tested and the graph relating frame current to frame-angle was a straight line parallel to the frame-angle axis.*

Suppose now that the wave-length is changed so that the frame is formatized in the square position, that is, the ratio of the frame-sides to the wave-length is given by the coordinates of one of the points A, e.g. A_1 in figure 1 (*a*). Then when the frame is oriented so that one side is parallel to the wave-front, the current measured by an instrument inserted in the middle of the side nearest the transmitter will be a maximum because the frame will be formatized. As the frame revolves in either direction it becomes more and more deformatized, and the current (still measured by putting the ammeter at the fixed point nearest the transmitter) will decrease until the diamond position is reached when the current will be a minimum. Thus a graph relating frame current with frame-angle should no longer be a straight line parallel to the frame-angle axis but should show a current peak whenever the frame is in the square position, and this will occur four times in each revolution.

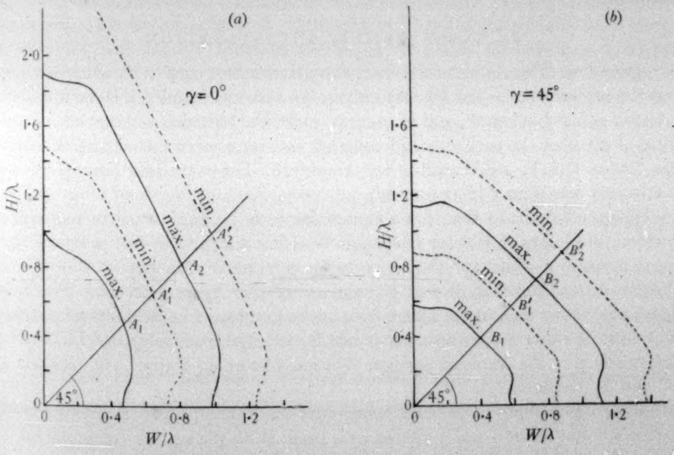


Figure 1. Graphical determination of theoretical formatizing dimensions for square receiving and transmitting frames.

By a precisely similar argument, when the wave-length is such that the frame is formatized in the diamond position, given by the coordinates of one of the points B, e.g. B_1 in figure 1 (*b*), then the graph relating frame current to frame-angle should show current-maxima whenever the frame is in the diamond position and current-minima whenever it is in the square position.

---

* See reference (2), figure 10, and also the last diagram in figure 4 of this paper.

Consider now the case in which the wave-length is such that the frame is deformatized in the square position; this is given by the coordinates of one of the points A', e.g. A_1' in figure 1 (*a*). Then since the frame is not completely deformatized in the diamond position, which would occur at B_1' in figure 1 (*b*), a graph relating frame current to frame-angle should show a minimum in the square position and a maximum in the diamond position. Similarly, a current-maximum should be obtained when the frame is in the square position if the wave-length be such that the frame is deformatized in the diamond position. It is to be expected that these current variations will be repeated when the wave-length is progressively decreased so that the conditions indicated by the coordinates of the points B_2, A_2, B_2' and A_2' are successively fulfilled; and so on for the points B_n, A_n, B_n' and A_n' in general.

I_S, I_D These theoretical deductions can best be shown on one graph by plotting against λ/W the logarithm of the ratio of the current I_S in the square position to that I_D in the diamond position. Positive peaks should then occur when the frame is formatized in the square position or deformatized in the diamond position, and negative peaks when it is formatized in the diamond position or deformatized in the square position.

§ 3. EXPERIMENTAL INVESTIGATION

Apparatus. The circuit of the short-wave transmitter used in these experiments was the ordinary series-fed Hartley circuit. To this was coupled a Hertzian dipole aerial of adjustable length, and in order to make the radiation as large and as uniform as possible the oscillator and coupling coil were screened with an iron box. An Osram D.E.T. 1.S.W. valve was employed. The wave-length range of the transmitter was from 5 to 12 metres.

Figure 1 (*a*) shows that, for a square frame to be formatized in the square position, the values of W/λ (or H/λ) must be either 0·47 or 0·82 (for points A_1 and A_2 respectively), whilst for the frame to be deformatized in the square position W/λ must be either 0·65 or 1·01 (for points A_1' and A_2' respectively). Similarly figure 1 (*b*) shows that for the square frame to be formatized in the diamond position W/λ must be either 0·39 or 0·72 (for points B_1 and B_2 respectively), and for it to be deformatized in the diamond position W/λ must be either 0·57 or 0·87 (for points B_1' and B_2' respectively).

As the wave-length range of the transmitter was limited, square receiving frames having sides of about 3 and 5 metres were used. With the smaller frame the minimum value of W/λ could be 3/12 or 0·25 whilst, with the larger frame, the maximum value of W/λ could be 5/5 or 1·0. In this way all the critical values deduced above from figure 1 could be studied.

One method of keeping the current measuring-instrument fixed in space as the frame revolved was as follows. The aerial wire of the frame was passed round insulated pulleys on the four corners of a suitable wooden framework which could rotate on a horizontal axis secured at a little distance from the top of a 40-foot scaffold pole. The current was measured by a vacuum thermojunction. Where the thermojunction was inserted in the frame, the aerial wire was shunted by a condenser of 0·0005-μF. capacity which tuned that portion of the frame to the wave-length in use. The current in the shunted portion of the aerial wire varied in the same way as that in the adjacent parts of the frame but was of greater magnitude and consequently easier to measure. The thermojunction was connected by a short length of flex to a microammeter which rested on a small platform hanging from a pivot clamped to the framework. The condenser, thermojunction and microammeter could be moved along one side of the frame and could be clamped in any required position.

The height of the frame above the ground was such that any change in its capacity to earth due to its revolving was negligibly small. The transmitter with a horizontal radiating dipole was placed vertically below the frame, so that if waves reflected from the ground affected the frame they were incident at the same angle as the waves arriving directly from the transmitter. The microammeter was read from the ground by means of a telescope and mirror because an observer near the frame would distort the field.

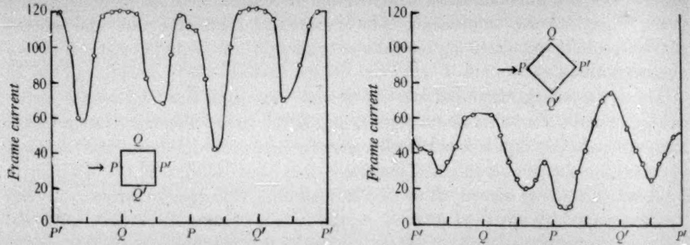
A second method of obtaining the desired current-measurements was to duplicate the thermojunction, microammeter and condenser. One thermojunction connected to a microammeter and a condenser was fixed in the middle of one of the frame sides, and another identical thermojunction, also connected to a microammeter and a condenser, was fixed at one corner of the frame. The whole frame with its two fixed instruments could revolve from the square to the diamond position and that instrument was read which was occupying the particular point in space at which the current was to be measured. In this way the ratio of the current when the frame was in the square position to the current when the frame was in the diamond position could be determined for a series of different wave-lengths.

Some of the observations were made with the frames in the horizontal plane about five feet above the ground and rotating about a vertical axis.

Current-measurements. It is convenient to consider the current-measurements under two sub-headings: (i) current-variations with a fixed frame, the instrument moving round the frame perimeter; and (ii) current-variations with a revolving frame, the instrument fixed in space relative to the transmitter.

(i) Frame fixed: instrument moving round frame. Experiments with the instrument moving round a fixed frame were carried out to test the form of the current-distribution round the frame when the latter was formatized and deformatized in both the square and the diamond positions. As an example the results with a frame 3·06 metres square are shown in figures 2 and 3 for wave-lengths of 6·52 and 7·89 metres respectively. For these wave-lengths W/λ is 0·47 and 0·39 respectively and hence the frame is formatized in the square positions and deformatized in the diamond positions for the case shown in figure 2, whilst it is deformatized in the square positions and formatized in the diamond positions for the case shown in figure 3. These results are discussed in § 4 below.

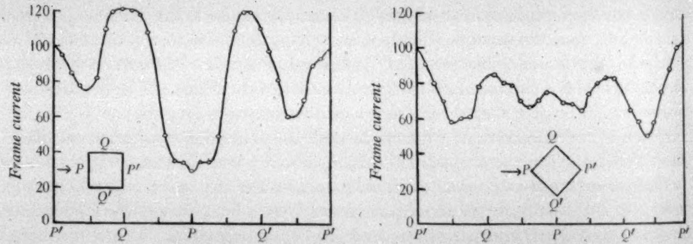
(ii) Frame revolving: instrument fixed in space. In the series of experiments with one measuring-instrument in the frame the procedure was as follows. The frame was turned to the diamond position and the instrument was placed at the corner of the frame nearest the transmitter. After the observation had been made through the telescope, the instrument was moved a short distance along the frame which was then revolved until the instrument came back to the same position in relation to the transmitter. The aerial wire moved round the insulated pulleys at the corners of the framework as the thermojunction etc., to which the ends of the aerial wire were attached, moved along the side of the frame.



(a)
Frame formatized

(b)
Frame deformatized

Figure 2. Current-distribution round a fixed frame with side measuring 3·06 m. Wave-length, 6·52 m.



(a)
Frame deformatized

(b)
Frame formatized

Figure 3. Current-distribution round a fixed frame with side measuring 3·06 m. Wave-length, 7·89 m.

In this way the frame was turned from one diamond position through a square position to the next diamond position. This procedure was adopted in the case of the 3-metre frame for wave-lengths varying from 5 to 12 metres, and for the

5-metre frame for wave-lengths varying from 5 to 10 metres. Some of the results are shown in the series of separate diagrams arranged in figures 4 and 5.

In the case of the frame with two fixed thermojunctions, the current measurements could only be taken when the frame was in the square or the diamond position. This was necessarily so because only when the frame was turned into one of those two positions was a thermojunction brought into that particular point in space at which the current was being measured. In the case of these experiments that particular point in space was the point of the frame nearest to the transmitter (points *P* on the inset diagrams of figures 2 and 3). The results of these experiments are plotted as circles in figure 6, in which the gain in the current I_S when the frame was in the square position over that I_D when the frame was in the diamond position is expressed in decibels. The reference letters on this figure correspond with those in figures 1 (*a*) and (*b*).

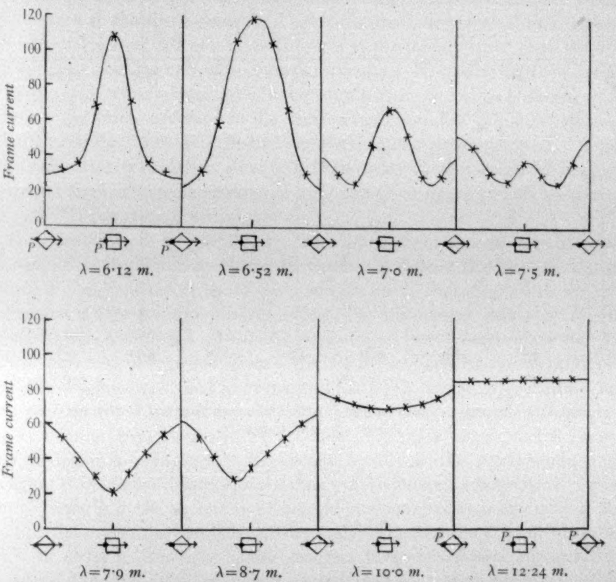


Figure 4. Current-measurements at *P* with frame revolving and instrument fixed. Frame 3·06 metres square.

All these results will now be considered in the light of the theoretical conclusions of § 2 above.

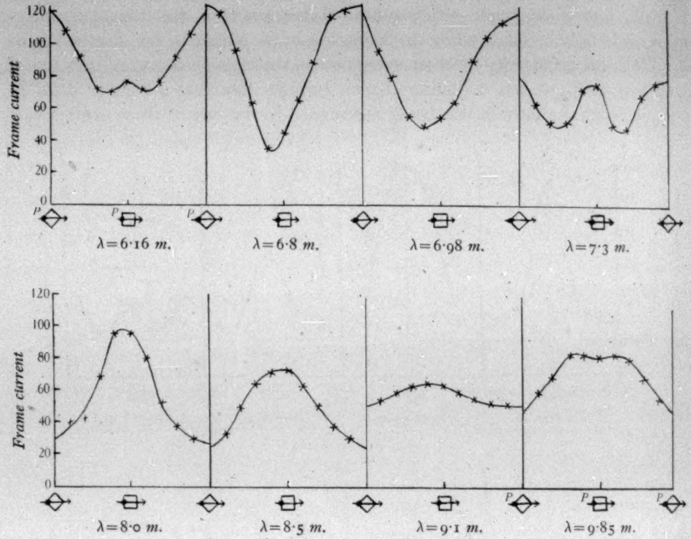


Figure 5. Current-measurements at P with frame revolving and instrument fixed. Frame 4·93 metres square.

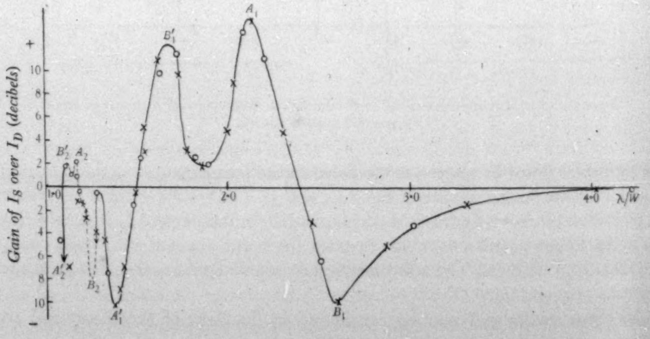


Figure 6. Variation of current-ratio with wave-length as a square frame revolves from the square to the diamond position.

§ 4. DISCUSSION OF RESULTS

From figures 2 and 3, which show the current-distribution round a fixed frame, the first conclusion that is evident is the one previously obtained; namely, the fixity of the current nodes and antinodes in space and their complete independence of the orientation of the frame with respect to the wave-front and to the position of the thermojunction and condenser.

The second conclusion arises from a comparative study of the temporal variation of the current at the fixed point P. This is the main objective of this paper. In figure 2 the current at this point has *decreased* as the frame turned from the square to the diamond position and so became deformatized, W/λ being 0·47 and not 0·39. The reverse has taken place with the current at P in figure 3. Here the current *increased* as the frame revolved to the diamond position, but W/λ now equals 0·39 for which ratio the frame is formatized in the diamond position.

Although these conclusions also apply to the other current antinodes in figure 2, they are not applicable to the antinodes at Q and Q' in figure 3. This is referred to again below in connection with figure 6.

Let us now turn to figures 4 and 5, which depict the current-variations at the fixed point P as the frame revolves. In figure 4 it is seen that there is a gradual transition throughout the range of wave-lengths used. Starting at 6·12 metres the variation due to formatizing is clearly indicated by the current peak when the frame is in the square position, and this becomes more pronounced when the wave-length is increased to 6·52 metres and the difference between the currents in the square and diamond positions is greatest. That is, the frame is formatized in the square position for a wave-length of 6·52 metres; this corresponds to point A_1, figure 1. Between 6·52 metres and 7·9 metres a transition occurs, and at about 7·9 metres the frame is formatized in the diamond position; this corresponds to point B_1 in figure 1. For wave-lengths longer than 7·9 the difference between the currents when the frame is in the square and diamond positions becomes less until at about 12 metres the current is sensibly constant, showing that there is no formatizing effect.* In order to illustrate in one diagram these progressive changes arising from points A_1 and B_1 in figure 1, and also the changes recorded in figure 5 which arise from points A_1' and B_1' in figure 1, together with other similar results not here figured, the ratio of the current I_S when the frame was in the square position to the current I_D when it was in the diamond position has been expressed in decibels, and the resultant values are marked as crosses in figure 6. With reference now to figure 6, it has been shown on page 392 that peaks should occur when the frame was either formatized in the square position (points A) or deformatized in the diamond position (points B') and troughs should occur when the frame was either formatized in the diamond position (points B) or deformatized in the square position (points A'). In the table on p. 398 the theoretical positions of these peaks (+) and troughs (−) are taken from figures 1 (*a*) and 1 (*b*) and are compared with the experimental values from figure 6.

---

* Cf. figure 10 on page 84 of reference (2).

A portion of the graph is shown dotted because, in order to obtain sufficient experimental values to fix the exact position of the curve, the change in λ/W had to be of the order of one per cent and this could not be controlled to the required degree of accuracy with the apparatus available. Nevertheless a slight change in wave-length showed at once whether any two adjacent points were on a portion of the curve with a positive or with a negative slope. Thus it is certain that where the value of λ/W is in the neighbourhood of 1·3 there are two troughs and not one, and where it is about 1·15 there are similarly two peaks.

Critical values of λ/W

Reference letter	B_1	A_1	B_1'	A_1'	B_2	A_2	B_2'	A_2'
Theoretical values from figure 1	−2·56	+2·12	+1·75	−1·54	−1·39	+1·22	+1·15	−0·99
Experimental values from figure 6	−2·58	+2·13	+1·69	−1·38	−1·25 (?)	+1·17 (?)	+1·11 (?)	−?

Hence it is thought that, in spite of the difficulties of these short-wave experiments, the agreement between the number and order of the peaks and troughs as determined by experiment and as deduced from figure 1 tends to support the theory of formatization upon which the graphs of figure 1 are based. This agreement is perhaps all the more interesting when it is realized that with the shorter wave-lengths there may be at least two current antinodes along any one limb of the frame, the current at any instant being in opposite phases at adjacent antinodes.

In conclusion it may be asked whether the current-increase which occurs when the frame becomes formatized is due to a decrease in the apparent impedance of the whole frame, or to a redistribution of a constant total current between the several current-antinodes. Figure 2 suggests that the whole frame has a reduced impedance when it is formatized, whilst figure 3 seems to support the second suggestion, since the current at Q and Q' seems to have decreased when the frame became formatized. Other experiments now in progress indicate that probably both these factors are involved.

Of the practical consequences of the foregoing work, two are immediately evident. (i) It depends on the wave-length whether it is best to orient a given square frame aerial in the square or in the diamond position, and (ii) a current-actuated or voltage-actuated instrument will function best when placed at certain specific points round the frame.

We may now summarize the results of this investigation by the following statement. *When a short-wave square frame aerial revolves in its own plane, then in addition to spatial current variations round the perimeter due to the formation of fixed current nodes and antinodes, there may also be temporal current variations at any fixed point due to the fact that the frame may become alternately formatized and deformatized as it revolves.*

§ 5. ACKNOWLEDGMENT

We should like to record our thanks to Mr Watson Watt who kindly loaned the short-wave transmitter, and also to Mr Denis Taylor for assistance in carrying out the experiments.

REFERENCES

(1) Palmer. *Proc. roy. Soc.* A, **136**, 193 (1932).
(2) Palmer, Taylor and Witty. *Proc. phys. Soc.* **46**, 76 (1934).
(3) Palmer and Taylor. *Proc. phys. Soc.* **46**, 62 (1934).

535.62

RAPID MATHEMATICAL METHODS FOR TRICHROMATIC COLORIMETRY

BY J. G. HOLMES, A.R.C.S., B.SC., D.I.C.
Research Department, Chance Brothers and Company, Ltd.

Received November 6, 1934. *Read in title March* 1, 1935.

ABSTRACT. Methods are suggested for reducing the labour involved in colorimetric transformations, with particular application to routine industrial measurements. The calibration of an instrument by measurement of illuminant *B* is replaced by measurement of a suitable stable filter and the illuminant in usual use. A nomographic method of conversion from instrument readings to the C.I.E. basis is suggested, and a graphical method is given for conversion to terms of hue and saturation. Scales are given for a graphical method of converting from the spectrophotometric curve of a coloured specimen to the trichromatic coefficients with a series of illuminants on the C.I.E. basis. The errors of these methods are estimated for comparison with the errors of observation on a Hilger-Guild trichromatic colorimeter.

§ 1. INTRODUCTION

THE majority of those who have investigated the use of a system of trichromatic colorimetry have developed some arithmetical or graphical method of simplifying the mathematics of the transformation with which they have been concerned. This has become increasingly necessary within the last fifteen years, owing to the continued use of different systems of primaries and to the need of coordination of the trichromatic system with the monochromatic-plus-white system. The basic principles of several transformations were given in two papers by Ives(1) and were followed by arithmetical applications by Froelich(2) in America and by Guild(3) in this country. Geometric methods and a variety of arithmetical methods have been derived since. The majority of these methods are arithmetically accurate; it is the purpose of this paper to put forward both graphical and observational methods of a sufficiently high order of accuracy for industrial or routine measurements, and for guidance in cases where the extreme accuracy required in the standardization of instruments or the designation of standard colours is not required. An accuracy of 1 per cent is all that can usually be expected in industrial colorimetry, and a mathematical tolerance of $\frac{1}{10}$ of this amount allows of compromises which effect a considerable acceleration of the calculations involved in some of these transformations.

§ 2. OBSERVATIONAL ERROR

The most recent measurements(4, 5) of the hue limen of fully saturated colours and of the saturation limen throughout the spectrum indicate that differences of hue of less than $2{\cdot}0\,\mathrm{m}\mu$., and that differences of saturation of less than 5 per cent, cannot be discerned in a single observation, except in certain limited areas of the colour triangle. For routine measurements, sufficient accuracy is assured if the gross probable error of the observational method is definitely less than the liminal tolerance at that point.

The errors of the method comprise the instrumental inaccuracies, the errors of reading the scales, and the errors of the transformation from scale readings to absolute values. The first of these may be sub-divided into mechanical inaccuracies which will be negligible in all well-designed instruments, and the inaccuracies introduced by irregular variation of the source of light and of the components of the optical system. These latter errors are extremely difficult to estimate but precautions may be taken to reduce their effect. The errors of reading depend on the accuracy and clarity of the scales, which again can usually be made fairly satisfactory. The errors of the transformations are the subject of this discussion.

Besides the errors inherent in the method there are the errors in the setting of the colorimeter, and an attempt has been made to estimate the error which obtains in the use of the Hilger-Guild trichromatic colorimeter. Statistical treatment has been applied to thirty-four routine measurements of illuminant *B* by two normal observers, the measurements being spread over several weeks and carried out with different solutions and different arrangements of the auxiliary apparatus. Each measurement was the mean of two settings taken with both the colorimeter lamp and the standard lamp maintained at steady potentials, and the readings of the three sectors on the instrument were reduced to unit sum for the purpose of comparison. The coefficients of variation for each observer were as shown in table 1.

Table 1. Observers' coefficients of variation (per cent)

Observer	Red sector	Green sector	Blue sector	Mean
A. H. A.	1·15	1·76	0·90	1·27
J. G. H.	1·10	1·25	0·66	0·99
Mean	1·12	1·50	0·78	1·13

The procedure which the author has adopted is that two settings are made on each specimen and the mean, obtained by mental arithmetic, is written down. This is taken as sufficiently reliable for most commercial purposes, but if greater accuracy is required the whole series of observations is repeated by the same method until the reliability of the results is ensured.

§ 3. ROUTINE CALIBRATION OF INSTRUMENT AND OBSERVER

The method to be employed for the calibration of the Hilger-Guild instrument is that in which a match is made on illuminant *B* and subsequent readings on any colour are divided by the sector readings of this match, so that the unit equation of illuminant *B* is $1/3R+1/3G+1/3B$. If the colorimeter is regularly used with an illuminant operated at 2360° K., as for signal glassware, or with illuminant *A*, as for lighting fittings, it is convenient to eliminate the preliminary setting on illuminant B, and this may be done by making a setting on the illuminant in use and applying a correction during the transformation of subsequent readings. Two methods are available for this.

It is frequently of value to know the unit equations of the standard illuminants in terms of the instrumental system. The original calibration of the instrument by spectrophotometric measurement of the filters gives the colour equations, in terms of the C.I.E. basis, of the unitary stimuli corrected for intensity, so that illuminant B is matched by numerically equal proportions of each, and these equations may be applied to transform the accepted C.I.E. colour equations of the illuminants to the instrumental basis. For the particular colorimeter which the author is considering the equations are as shown in table 2.

Table 2. Calibration equations for Hilger-Guild colorimeter No. H93304/27206

R, X, Y, Z
G
B
Q

Instrumental stimuli	$R=0·744X+0·288Y+0·000Z$ $G=0·151X+0·721Y+0·075Z$ $B=0·150X+0·046Y+0·825Z$
Illuminant A	$Q_A=0·4476X+0·4074Y+0·1450Z$ $=0·501R+0·356G+0·143B$
Illuminant B	$Q_B=0·3484X+0·3516Y+0·3000Z$ $=0·333R+0·333G+0·333B$
Illuminant C	$Q_C=0·3100X+0·3162Y+0·3738Z$ $=0·269R+0·305G+0·426B$
Equi-energy white	$Q_E=0·3333X+0·3333Y+0·3333Z$ $=0·309R+0·315G+0·376B$
2360° K.	$Q_{2360^\circ}=0·4894X+0·4150Y+0·0956Z$ $=0·575R+0·340G+0·085B$
1900° K.	$Q_{1900^\circ}=0·5372X+0·4114Y+0·0514Z$ $=0·657R+0·309G+0·034B$

If a match is made on one of these illuminants and the sector readings are divided by factors chosen to yield the appropriate unit equation on the instrumental system, the same factors can be applied to subsequent sector readings to reduce the results for all colours to the instrumental system. This method of eliminating illuminant B is not of high accuracy in the blue when used with a source at 2360° K. A higher accuracy is obtainable by using a glass or stable gelatine filter which, when combined with the usual illuminant, gives a colour similar to that of illuminant B, which is itself a combination of a liquid filter and illuminant A. This gives a calibration as accurate as illuminant B itself would give, and has the advantage that the instrument may be calibrated without the use of fresh solutions in the Davis-Gibson filter or without any doubts as to the freshness of the solutions. It is advisable to have the spectrophotometric curve of the glass filter determined accurately and to calculate the equation of the colour which it gives with the illuminant with which it is most used. When a match is made on this combination the factors necessary to reduce the sector readings to the unit equation of the colour can be used for a conversion of subsequent sector-readings. For example, a particular glass of the ordinary half-watt-to-daylight type, when used with illuminant *A*, gives an equation as follows:

$$Q_D = 0{\cdot}270R + 0{\cdot}378G + 0{\cdot}352B.$$

Q_D

This glass has been used consistently without any apparent error in the author's colorimeter. It is possible to select other blue glasses which would give a closer approximation to illuminant *B* with either 2360° K. or illuminant *A*.

§4. CONVERSION FROM THE INSTRUMENTAL SYSTEM TO C.I.E. SYSTEM

The coefficients of the colour equation of any particular specimen on the instrumental system may be converted to the C.I.E. system by replacing the instrumental stimuli, as they appear in the equation, by their own equations on the C.I.E. system, which will have been obtained in the original calibration of the instrument.

The arithmetical transformation is easy, but a nomographical method is available which converts the unit equation on the instrumental system to an equation on the C.I.E. system, the sum of whose coefficients is sufficiently close to unity to enable the unit equation to be written down by the aid of a slide rule. A nomogram made on the principles outlined below, on a card measuring 20 in. by 30 in., gives figures reliable to 0·001 in the unit equation.

Suppose the colour to be represented by

$$C = r.R + g.G + b.B,$$

r, g, b

where r, g, b are known, and

$$r + g + b = 1.$$

The original calibration of the colorimeter would yield equations of the form

$$R = \rho_1.X + \rho_2.Y + \rho_3.Z,$$

ρ

$$G = \gamma_1.X + \gamma_2.Y + \gamma_3.Z,$$

γ

$$B = \beta_1.X + \beta_2.Y + \beta_3.Z.$$

β

The required equation is then

$$C = x.X + y.Y + z.Z,$$

where

$$x = r.(\rho_1 - \gamma_1) + \gamma_1 + b.(\beta_1 - \gamma_1),$$

x

$$y = r.(\rho_2 - \gamma_2) + \gamma_2 + b.(\beta_2 - \gamma_2),$$

y

$$z = r.(\rho_3 - \gamma_3) + \gamma_3 + b.(\beta_3 - \gamma_3).$$

z

In a nomogram of the form shown in figure 1 the *R*, *G* and *B* scales will be at equal separation, the *G* scale being reversed and with half the unit of the *R* and *B* scales, so that $r + g + b = 1$. If the total separation of the *R* and *B* scales is *D*, and D

d the unit of these scales is *d*, the positions and units of the *X*, *Y* and *Z* scales can be shown to be as follows:

Position of *X* scale ... $\dfrac{(\beta_1-\gamma_1).D}{\rho_1+\beta_1-2\gamma_1}$ from *R* scale.
Position of zero on *X* scale $-\gamma_1$.
Unit distance on *X* scale $d/(\rho_1+\beta_1-2\gamma_1)$.
Position of *Y* scale ... $\dfrac{(\beta_2-\gamma_2).D}{\rho_2+\beta_2-2\gamma_2}$ from *R* scale.
Position of zero on *Y* scale $-\gamma_2$.
Unit distance on *Y* scale $d/(\rho_2+\beta_2-2\gamma_2)$.
Position of *Z* scale ... $\dfrac{(\beta_3-\gamma_3).D}{\rho_3+\beta_3-2\gamma_3}$ from *R* scale.
Position of zero on *Z* scale $-\gamma_3$.
Unit distance on *Z* scale $d/(\rho_3+\beta_3-2\gamma_3)$.

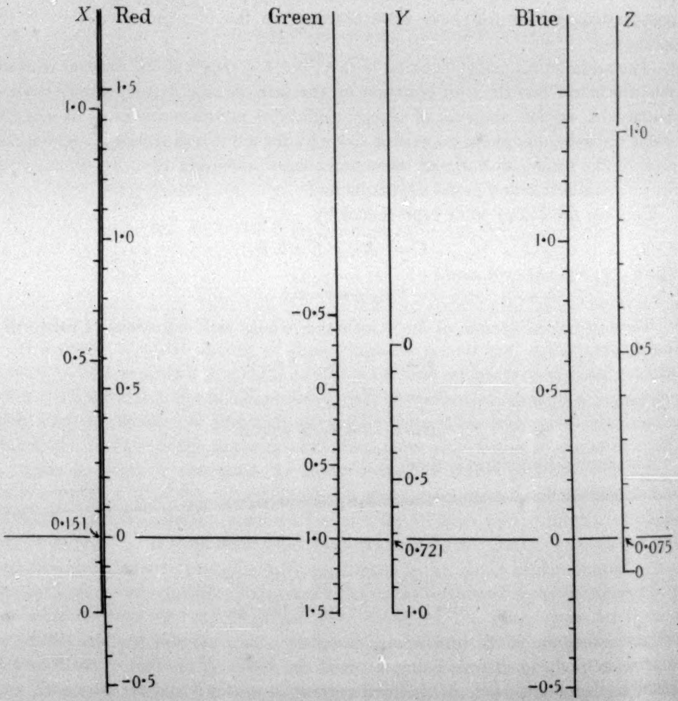

X
Red
Green
Y
Blue
Z
1·5
1·0
1·0
1·0
-0·5
1·0
0·5
0
0·5
0·5
0·5
0
0·5
0·5
0·151
0
1·0
0·721
0
0·075
0
0
1·5
1·0
-0·5
-0·5

Figure 1. Nomogram for conversion from instrument basis to C.I.E. basis (H 93304/27206).

The nomogram shown is for the Hilger-Guild colorimeter, whose calibration equations are given in table 2.

As an instance of the use of this nomogram, consider the conversion of the colour equation of a half-watt-to-daylight glass, with illuminant A, from the instrumental basis to the C.I.E. basis. The unit equation on the former basis is

$$Q_D = 0{\cdot}270R + 0{\cdot}378G + 0{\cdot}352B.$$

If a straight line, such as a stretched steel wire, be placed across the diagram through the points $0{\cdot}270R$ and $0{\cdot}352B$, it will also be found to pass through the point $0{\cdot}378G$, and so this line now represents the equation of the colour. This line will be found to intersect the X, Y and Z scales at the points $0{\cdot}311X$, $0{\cdot}366Y$ and $0{\cdot}318Z$. The sum of these coefficients is 0·995 and the unit equation may be written down at once:

$$Q_D = 0{\cdot}312X + 0{\cdot}368Y + 0{\cdot}320Z.$$

These coefficients are the same as those found by the arithmetical transformation. The same process may be applied if some of the coefficients of R, G and B in the unit equation are negative or greater than unity. The sum of the values of X, Y and Z given by the intersections will always be close to unity, so the evaluation of the unit equation on the C.I.E. basis is a matter of mental arithmetic or a simple slide-rule conversion.

§ 5. COLORIMETRIC SATURATION

The specification of a colour in terms of the coefficients of the three unitary stimuli X, Y and Z is a complete description which is internationally comparable, but it is not generally possible to visualize the colour immediately from those figures, and although the specification is undoubtedly the best for all records or exact statements, there is a use for the conception of hue and saturation as a means of describing a colour. This latter method is liable to much abuse, for it can only be exact when care is taken to specify the white light which is taken as the reference point of the hue and saturation, and different sets of figures are not comparable unless the same white diluent is used for each.

The method is based on the experimental fact that any colour may be matched by the combination of two stimuli in suitable proportion, provided that the three points given by plotting their trichromatic coefficients on a linear scale are collinear. In general, the two stimuli chosen are the equi-energy white, which is termed the diluent, and a spectral colour, which is termed the hue wave-length. The colorimetric saturation is then the percentage of the spectral component when both stimuli are measured in units of the same luminosity.

The transformation from the trichromatic system to the homogeneous-plus-heterogeneous system may be performed in various ways. Two exact methods have been proposed, one by Smith and Guild(6) and one by Judd(7), both of which are assisted if a plot of the spectrum locus is used. In practice the system is not used under conditions where exact specification is necessary, and an approximate transformation can be read off a graph if loci of equal saturation are plotted inside the spectrum locus. It is obvious that these loci will only apply to one particular white diluent, but as it is not the general practice to use anything but equi-energy white for this purpose, the iso-saturation loci with reference to equi-energy white may be calculated with advantage.

Taking the general case, let us define the following stimuli on some trichromatic system:

R, G, B	Coefficients of heterogeneous stimulus (white diluent)	R, G, B.
r', g', b'	Coefficients of homogeneous stimulus (hue wave-length or spectral component)	r', g', b'.
r, g, b	Coefficients of combined colour of above, in specified proportions	r, g, b.
L_w, L_λ, L_c	Luminosities of above stimuli, respectively ...	L_w, L_λ, L_c.
L_r, L_g, L_b	Luminosities of unitary stimuli of system ...	L_r, L_g, L_b.
σ	Saturation of combined colour	σ.

One luminous unit of the combined colour is proportional to the reciprocal of its luminosity:

$$\frac{1}{L_c}=\frac{1}{r.L_r+g.L_g+b.L_b} \qquad \text{......(1).}$$

Also, by definition of "saturation",

$$\frac{1}{L_c}=\frac{\sigma}{L_\lambda}+\frac{1-\sigma}{L_w} \qquad \text{......(2).}$$

Because the three points are collinear:

$$\frac{r-R}{r'-R}=\frac{g-G}{g'-G}=\frac{b-B}{b'-B}=\delta,$$

or

$$r=R+\delta\,(r'-R), \text{ etc.}$$

After insertion of these in equation (1),

$$\frac{1}{L_c}=\frac{1}{L_w+\delta\,(L_\lambda-L_w)} \qquad \text{......(3).}$$

Then, from equations (2) and (3),

$$\frac{1}{\delta}=1+\frac{L_\lambda}{L_w}\left(\frac{1}{\sigma}-1\right).$$

d_λ d_σ Let the linear distance from the diluent to the hue wave-length on a chart such as figure 2 be d_λ, and let the distance from the diluent to the combined colour be d_σ.

Then

$$\begin{aligned}\frac{d_\sigma}{d_\lambda}&=\frac{r'-R}{r-R}\\&=\frac{1}{\delta}\\&=1+\left(\frac{1}{\sigma}-1\right)\frac{r'.L_r+g'.L_g+b'.L_b}{R.L_r+G.L_g+B.L_b}\end{aligned} \qquad \text{......(4).}$$

The final equation gives the ratio of the distances on the colour chart between points corresponding to the diluent, the spectral component and the combined colour, in terms of the known constants of the system. Reducing this to the C.I.E. system, we have

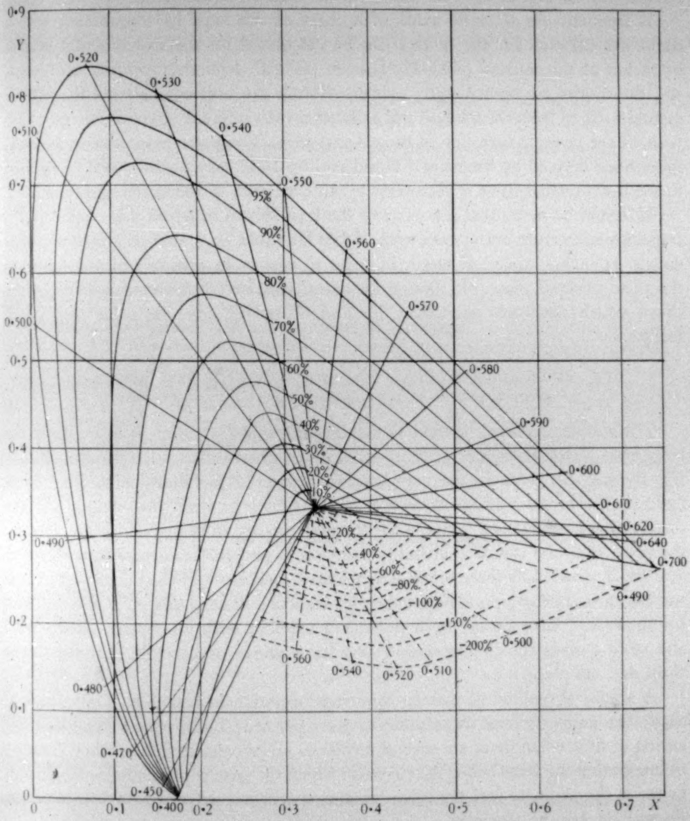


Figure 2. Iso-saturation loci on C.I.E. basis.

$$\frac{d_\sigma}{d_\lambda} = 1 + \left(\frac{1}{\sigma} - 1\right) \cdot 3y,$$

which is the same as the equation obtained in Smith and Guild's method.

The use of the distances of the points on the chart from the diluent permits of a rapid calculation of the iso-saturation loci, and the curves shown in figure 2 were obtained by this method. From these curves, tables may be compiled for permanent record, but the results are best used in their diagrammatic form.

If frequent use is to be made of a chart of this type in conjunction with a particular colorimeter, the figures can be calculated on the instrumental system by means of the relation given as equation (4) with equi-energy white as diluent, and the results are then directly converted from the unit equation on the instrumental basis to terms of the hue and saturation with reference to equi-energy white, without the intermediate step of the C.I.E. system. The more exact algebraic transformations derived by Smith and Guild and by Judd may equally well be applied to the instrumental basis if necessary, by an extension of the same reasoning.

It should be noted that if a chart of the type shown in figure 2 is employed for transformation from one system to the other it should be drawn on a card approximately 24 inches square or larger in order to ensure an accuracy of one figure in the third decimal place. If this is not practicable or convenient, sections of the chart, which need not necessarily include the centre, can be used with equal facility.

§ 6. TRANSFORMATION FROM SPECTROPHOTOMETRIC DATA TO TRICHROMATIC COEFFICIENTS

$E\bar{x}$, $E\bar{y}$, $E\bar{z}$
E
$\bar{x}$, $\bar{y}$, $\bar{z}$

The simple algebraic method of making this transformation is to multiply the brightness factor at each of a series of wave-lengths by the corresponding figures $E\bar{x}$, $E\bar{y}$ and $E\bar{z}$ given in the computational table of the illuminant. This latter figure is the product of the energy E of the source at that wave-length and the distribution coefficients $\bar{x}$, $\bar{y}$ and $\bar{z}$ for the equi-energy stimulus, which are actually the product of the standard visibility function V and the trichromatic coefficients x, y and z, divided by the coefficient y, which is a measure of the luminosity at that wave-length. The computational tables for standard illuminants A, B and C(6) and for 1900° K.(8) have already been published together with illustrative calculations, and table 3 gives the computational data for a source at 2360° K., carried to the third decimal place.

A graphical method of making this transformation is available. Difficulties of accurate planimetry limit its accuracy to $\pm 0{\cdot}5$ per cent, but it effects a considerable saving of time when there are several results to be converted. The method consists of integrating the areas below the curves in which the spectrophotometric brightness factor is the ordinate and the wave-lengths, spaced according to the values of $E\bar{x}$, $E\bar{y}$ and $E\bar{z}$, are the abscissae.

A different wave-length scale will be required for each unitary stimulus and for each illuminant, but when once the charts have been prepared on stiff paper they may be used almost indefinitely. The spacing of the wave-length scales is such that the intervals are proportional to the value of $E\bar{x}$, $E\bar{y}$ or $E\bar{z}$ at corresponding wave-lengths and table 4 gives the scales for the standard illuminants A, B and C and for Planckian sources at 2360° K. and 1900° K. on a basis of a 10-in. scale for the coefficient of Y, giving points at every $10\,\mathrm{m}\mu$.

The integrated visible-brightness factor of the specimen is given by the value which the coefficient of Y has before reduction to the unit equation has taken place.

§ 7. ACKNOWLEDGMENT

The author wishes to thank the Directors of Messrs Chance Brothers and Co., Ltd. for permission to publish these notes, and Dr W. M. Hampton, the Director of Research, whose interest and advice have been of great value.

REFERENCES

(1) Ives, H. E. "Transformation of colour mixture equations from one system to another." *J. Franklin Inst.* **180**, 673 (1915) and **195**, 23 (1923).

(2) Froelich, C. L. "Algebraic method for the calculation of colour mixture transformation diagrams." *J. opt. Soc. Amer.* **9**, 31 (1924).

(3) Guild, J. "The transformation of trichromatic mixture data: algebraic methods." *Trans. opt. Soc.* **26**, 95 (1924–5). "Geometrical solution of colour mixture problems." **26**, 139 (1924–5).

(4) Martin, J. C., Warburton, F. L. and Morgan, W. J. "Determination of the sensitiveness of the eye to differences in the saturation of colours." *Spec. Rep. Ser. med. Res. Coun.*, Lond., No. 188 (1933).

(5) Wright, W. D. and Pitt, F. G. H. "Hue discrimination in normal colour vision." *Proc. phys. Soc.* **46**, 459 (1934).

(6) Smith, T. and Guild, J. "c.i.e. colorimetric standards and their use." *Trans. opt. Soc.* **33**, 73 (1931–2).

(7) Judd, D. B. "1931 i.c.i. standard observer and coordinate system for colorimetry." *J. opt. Soc. Amer.* **23**, 359 (1933).

(8) Gibson, K. S. and others. "Specification of railway colours." *Proc. Amer. Railr. Ass.* (Signal Section), **30**, 385 (1933).

Table 3. Computational data for 2360° K. illuminant (C_2=14,350 micron-degrees)

Wave-length (mμ.)	Relative energy distribution	$E\bar{x}$	$E\bar{y}$	$E\bar{z}$
380	0·38708	0·001	0·000	0·002
390	0·51039	0·002	0·000	0·010
400	0·66420	0·009	0·000	0·042
410	0·85016	0·035	0·001	0·165
420	1·0485	0·132	0·004	0·635
430	1·3355	0·355	0·015	1·735
440	1·6427	0·537	0·035	2·694
450	1·9950	0·630	0·071	3·316
460	2·3977	0·654	0·135	3·762
470	2·8530	0·523	0·243	3·442
480	3·3620	0·302	0·438	2·564
490	3·9283	0·118	0·766	1·713
500	4·5504	0·021	1·379	1·161
510	5·2318	0·046	2·469	0·776
520	5·9704	0·355	3·974	0·438
530	6·7671	1·051	5·472	0·268
540	7·6234	2·078	6·820	0·145
550	8·5349	3·471	7·964	0·070
560	9·5028	5·300	8·864	0·035
570	10·521	7·521	9·410	0·021
580	11·594	9·962	9·465	0·018
590	12·713	12·245	9·024	0·013
600	13·879	13·840	8·218	0·010
610	15·085	14·200	7·122	0·004
620	16·336	13·092	5·845	0·003
630	17·620	10·615	4·381	0·000
640	18·937	7·962	3·110	0·000
650	20·279	5·395	2·035	0·000
660	21·652	3·348	1·239	0·000
670	23·044	1·890	0·692	0·000
680	24·454	1·074	0·391	0·000
690	25·879	0·551	0·199	0·000
700	27·309	0·292	0·105	0·000
710	28·754	0·156	0·057	0·000
720	30·182	0·082	0·028	0·000
730	31·645	0·041	0·015	0·000
740	33·079	0·022	0·009	0·000
750	34·522	0·010	0·003	0·000
760	35·851	0·007	0·002	0·000
770	37·357	0·000	0·000	0·000
780	38·748	0·000	0·000	0·000
		117·925	100·000	23·042

$$Q_{2360^\circ\,\mathrm{K.}} = 0{\cdot}4894X + 0{\cdot}4150Y + 0{\cdot}0956Z.$$

Table 4. Abscissae for standard illuminants

Wave-length (mμ.)	Illuminant *A*			Illuminant *B*			Illuminant *C*		
	X	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*
380	0·000	0·000	0·000	0·000	0·000	0·000	0·000	0·000	0·001
390	0·000	0·000	0·001	0·001	0·000	0·004	0·001	0·000	0·003
400	0·001	0·000	0·006	0·004	0·000	0·018	0·005	0·000	0·027
410	0·005	0·000	0·026	0·016	0·000	0·075	0·024	0·001	0·113
420	0·021	0·001	0·099	0·063	0·002	0·301	0·095	0·003	0·457
430	0·066	0·002	0·319	0·202	0·007	0·980	0·318	0·010	1·488
440	0·146	0·006	0·705	0·441	0·020	2·158	0·664	0·029	3·249
450	0·245	0·015	1·222	0·716	0·044	3·568	1·063	0·064	5·298
460	0·348	0·031	1·790	0·978	0·084	5·005	1·430	0·120	7·308
470	0·440	0·059	2·350	1·190	0·148	6·298	1·716	0·207	9·050
480	0·500	0·118	2·788	1·319	0·250	7·233	1·884	0·340	10·268
490	0·529	0·189	3·083	1·375	0·407	7·808	1·954	0·537	10·992
500	0·537	0·329	3·272	1·390	0·647	8·138	1·973	0·823	11·390
510	0·540	0·569	3·396	1·394	1·181	8·327	1·979	1·233	11·606
520	0·561	0·960	3·469	1·422	1·529	8·426	2·008	1·798	11·713
530	0·641	1·516	3·510	1·519	2·214	8·477	2·110	2·518	11·767
540	0·815	2·214	3·533	1·725	3·037	8·504	2·324	3·375	11·795
550	1·115	3·024	3·545	2·067	3·963	8·518	2·676	4·328	11·809
560	1·575	3·916	3·550	2·567	4·936	8·524	3·184	5·316	11·815
570	2·228	4·853	3·553	3·234	5·895	8·526	3·846	6·269	11·818
580	3·093	5·790	3·555	4·052	6·785	8·528	4·635	7·128	11·820
590	4·162	6·681	3·556	4·984	7·563	8·529	5·509	7·859	11·821
600	5·382	7·487	3·557	5·973	8·217	8·530	6·410	8·345	11·822
610	6·659	8·183	3·558	6·956	8·749	8·531	7·278	8·929	11·822
620	7·869	8·755	3·558	7·381	9·167	8·531	8·052	9·295	11·822
630	8·891	9·193	3·558	8·650	9·475	8·531	8·672	9·562	11·822
640	9·667	9·505	3·558	9·077	9·687	8·531	9·120	9·742	11·822
650	10·209	9·714	3·558	9·434	9·825	8·531	9·421	9·857	11·822
660	10·554	9·843	3·558	9·656	9·907	8·531	9·608	9·926	11·822
670	10·754	9·916	3·558	9·780	9·953	8·531	9·705	9·963	11·822
680	10·864	9·957	3·558	9·845	9·977	8·531	9·758	9·982	11·822
690	10·923	9·978	3·558	9·878	9·988	8·531	9·784	9·992	11·822
700	10·953	9·989	3·558	9·894	9·995	8·531	9·796	9·996	11·822
710	10·969	9·995	3·558	9·902	9·997	8·531	9·802	9·998	11·822
720	10·977	9·997	3·558	9·905	9·999	8·531	9·805	9·999	11·822
730	10·981	9·998	3·558	9·908	9·999	8·531	9·806	10·000	11·822
740	10·983	9·999	3·558	9·909	10·000	8·531	9·807	10·000	11·822
750	10·984	10·000	3·558	9·909	10·000	8·531	9·807	10·000	11·822
760	10·984	10·000	3·558	9·909	10·000	8·531	9·807	10·000	11·822
770	10·985	10·000	3·558	9·909	10·000	8·531	9·807	10·000	11·822
780	10·985	10·000	3·558	9·909	10·000	8·531	9·807	10·000	11·822

Table 4 (*cont.*). Abscissae for standard illuminants

Wave-length (mμ.)	2360° K.			1900° K.		
	X	*Y*	*Z*	*X*	*Y*	*Z*
380	0·000	0·000	0·000	0·000	0·000	0·000
390	0·000	0·000	0·001	0·000	0·000	0·000
400	0·001	0·000	0·005	0·000	0·000	0·002
410	0·005	0·000	0·022	0·002	0·000	0·008
420	0·018	0·000	0·085	0·007	0·000	0·032
430	0·053	0·002	0·259	0·022	0·001	0·105
440	0·107	0·006	0·528	0·046	0·002	0·226
450	0·170	0·013	0·860	0·076	0·006	0·386
460	0·236	0·026	1·236	0·110	0·013	0·582
470	0·288	0·050	1·580	0·139	0·026	0·773
480	0·318	0·094	1·837	0·157	0·052	0·926
490	0·330	0·171	2·008	0·165	0·101	1·034
500	0·332	0·309	2·124	0·166	0·193	1·112
510	0·336	0·556	2·202	0·169	0·369	1·167
520	0·372	0·953	2·246	0·196	0·668	1·200
530	0·477	1·500	2·272	0·279	1·102	1·221
540	0·685	2·182	2·287	0·452	1·671	1·234
550	1·032	2·979	2·294	0·757	2·370	1·240
560	1·562	3·865	2·297	1·245	3·186	1·243
570	2·314	4·806	2·299	1·969	4·092	1·245
580	3·310	5·752	2·301	2·974	5·045	1·247
590	4·535	6·655	2·302	4·262	5·995	1·248
600	5·919	7·477	2·303	5·778	6·896	1·249
610	7·339	8·189	2·304	7·399	7·709	1·250
620	8·648	8·773	2·304	8·953	8·402	1·250
630	9·710	9·211	2·304	10·262	8·942	1·250
640	10·506	9·522	2·304	11·280	9·340	1·250
650	11·045	9·726	2·304	11·995	9·610	1·250
660	11·380	9·850	2·304	12·454	9·779	1·250
670	11·569	9·919	2·304	12·721	9·877	1·250
680	11·676	9·958	2·304	12·878	9·934	1·250
690	11·732	9·978	2·304	12·962	9·964	1·250
700	11·761	9·988	2·304	13·007	9·981	1·250
710	11·776	9·994	2·304	13·032	9·990	1·250
720	11·784	9·997	2·304	13·046	9·995	1·250
730	11·789	9·998	2·304	13·053	9·997	1·250
740	11·791	9·999	2·304	13·057	9·999	1·250
750	11·792	10·000	2·304	13·058	9·999	1·250
760	11·792	10·000	2·304	13·059	10·000	1·250
770	11·792	10·000	2·304	13·060	10·000	1·250
780	11·792	10·000	2·304	13·060	10·000	1·250

535·337:546.17

NOTE ON ADDITIONAL EXPERIMENTS ON THE EFFECTIVE ROTATION TEMPERATURE OF THE NEGATIVE GLOW IN NITROGEN

By N. THOMPSON, Ph.D.

Received November 14, 1934. *Read in title March* 1, 1935.

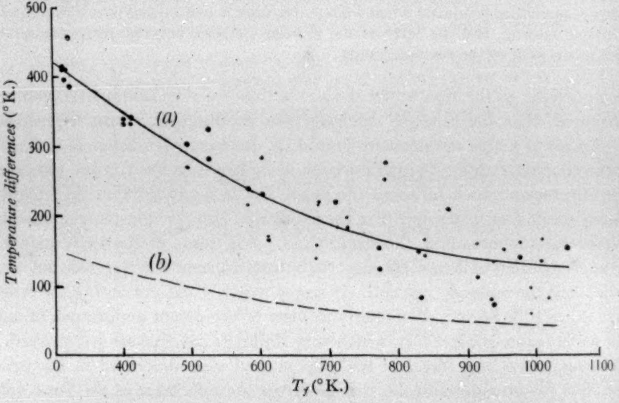
ABSTRACT. A summary is given of the results of further determinations of the effective rotation temperature produced when a discharge tube is maintained at a high temperature by external heating, and the form of the relation between the two temperatures is discussed in the light of the revised results.

In a previous publication under the above title(1) certain anomalous results were reported when the nitrogen discharge was produced in a tube surrounded by a furnace at a high temperature (700° C.). In these circumstances it was found that the effective rotation temperature could be less than the furnace temperature. In the discussion which followed the paper it was suggested that this might be a spurious result due to the use of a thermocouple (copper-constantan) unsuited to the temperatures measured. To decide this point, some of the experiments have since been repeated with new couples; the results indicate that the criticism was well founded and the anomaly not real. It seems worth while presenting the amended results in detail however, particularly in view of the recent appearance of another paper on the same subject(2) in which very different conclusions are reached.

The apparatus was identical with the second type described in the previous paper. The gas pressure and the voltage across the tube were of the same order as formerly, being respectively 1·2 mm. ± 0·4 per cent, and 384 V. ± 0·3 per cent, while the current was higher at 30 mA. The temperatures inside the tube were measured with either a tungsten-constantan or a nickel-nichrome thermocouple. The former was calibrated both before and after use (the two calibrations agreeing very well) and the latter after use only. The two instruments gave quite consistent results. It should perhaps be mentioned that the thermocouple reading was taken in every case while the discharge was running. This reading was higher than that which obtained in the absence of discharge, by an amount which decreased with increasing temperature and was about 30° C. at room temperature. The effective rotation temperatures were evaluated as before.

Several independent series of results were obtained, and they are summarized in the appended figure which shows $(\tau - T_f)$ plotted against T_f, where τ is the effective rotation temperature and T_f is the furnace temperature. In agreement with the previous results, this quantity falls with increasing temperature, but the earlier conclusion that it eventually becomes negative is not confirmed. The shape of the smoothed curve is essentially the same as that showing the temperature T_e

of the cathode in figure 5 of the earlier paper. The cathode temperature was not measured in these later experiments, but the earlier results may be taken as indicating roughly the general run of the curves, since both T_c and T_f depended ultimately on the same faulty thermocouple calibration. On this basis the probable shape of the curve showing $(T_c - T_f)$ against T_f is indicated in the figure by the dotted line, which is taken from figure 5 of the original paper. Although the two sets of results are not quantitatively comparable, there is no doubt whatever that the effective rotation temperature is higher than the temperature of the cathode at low temperatures (compare figures 3 and 5 of original paper) and it is also very probable that it always remains the higher of the two, as shown in the accompanying figure.



Curves showing (*a*) $(\tau - T_f)$ and (*b*) $(T_c - T_f)$ plotted against T_f.

If we continue, in default of any better estimate, to take T_c as an indication of the true gas temperature, these results show that the effective rotation temperature is higher than the gas temperature, as was found by Duffendack. On the other hand, that author's conclusion that the former temperature is independent of the latter is certainly not borne out. In the same paper he suggests that the difference $(\tau - T_c)$ may be due to an increase in the mean rotational kinetic energy of the molecule brought about directly during the exciting collision process. This hypothesis, however, does not seem able to account for the marked variation of this quantity with temperature, which is shown by the above graph.

It is a pleasure to express my thanks to Prof. Milner of Sheffield University for providing the facilities which enabled me to carry out these later experiments.

REFERENCES

(1) THOMPSON. *Proc. phys. Soc.* **46**, 436 (1934).
(2) DUFFENDACK, REVANS and ROY. *Phys. Rev.* **45**, 807 (1934).

535·338.4:546.45-31

THE BAND SPECTRUM OF BERYLLIUM MONOXIDE

By A. HARVEY, Ph.D., F.Inst.P., and H. BELL, M.A.
Physics Department, the Victoria University of Manchester

Received December 14, 1934. *Read in title March* 1, 1935.

ABSTRACT. It is shown that certain band-heads previously unaccounted for in the ultra-violet spectrum of BeO can be fitted by means of the equation

$$\nu_h = 29753{\cdot}0 + 1083{\cdot}1\,(v'+\tfrac{1}{2}) - 11{\cdot}1\,(v'+\tfrac{1}{2})^2 - 1148{\cdot}7\,(v''+\tfrac{1}{2}) + 11{\cdot}6\,(v''+\tfrac{1}{2})^2.$$

It is suggested that the final level of these bands is the same as the initial level of the near infra-red bands (i.e. a $^1\Pi$ level) and it is also suggested that both the ultra-violet BeO systems are of $^1\Pi \rightarrow {}^1\Pi$ type.

§ 1. INTRODUCTION

Excitation of the BeO molecule gives rise to a number of band systems. Two of these have been analysed in detail and have been shown to be $^1\Sigma \rightarrow {}^1\Sigma$ and $^1\Pi \rightarrow {}^1\Sigma$ respectively, lying in the blue-green in the first instance(1, 2) and in the near infra-red in the second(3). Band emission has also been observed in the near ultra-violet. Jevons(4) was the first to observe this but made no analysis; Bengtsson (*loc. cit.*), however, succeeded in making a vibrational analysis which accounted for some of the bands, although quite a number were not included. The purpose of this note is to indicate how these remaining bands may be accounted for.

§ 2. EXPERIMENTAL DETAILS

For the analysis of the $^1\Sigma \rightarrow {}^1\Sigma$ bands(1, 2) the spectrum had been excited by the burning of a beryllium salt in the carbon arc. In the present work a number of spectrograms were obtained in this way but the long exposures, coupled with the fact that such an arc requires constant attention, led us to make use of the metallic beryllium now obtainable and to run an uncondensed spark between two electrodes of the metal. Such a spark will run unattended for many hours. (The arc between electrodes of metallic beryllium was used by L. Herzberg(3) and was found to require constant attention.) The spectrograms were obtained in the first and second orders of the 21-ft. grating now mounted in this department. Comparison spectra from an iron arc were put on the plates by means of a set of shutters controlled from outside the grating room itself.

§ 3. THE ULTRA-VIOLET BAND SYSTEMS

Bengtsson succeeded in describing a number of the bands by means of the equation:

$$\nu_h = 31899 + 1006n' - 10n'^2 - 1126n'' + 10n''^2 \qquad \text{.........(1).}$$

Rewriting this in terms of the new quantum theory and the accepted notation ([5], p. 26), we get

$$\nu_h = 31959 + 1016\,(v'+\tfrac{1}{2}) - 10\,(v'+\tfrac{1}{2})^2 - 1136\,(v''+\tfrac{1}{2}) + 10\,(v''+\tfrac{1}{2})^2 \qquad \text{......(2).}$$

At the time when this analysis was made only the blue-green $^1\Sigma \rightarrow {}^1\Sigma$ bands had been analysed and it was clear that this new ultra-violet system was unrelated to the $^1\Sigma \rightarrow {}^1\Sigma$ system. The analysis of the near infra-red system by L. Herzberg, however, showed that these were of $^1\Pi \rightarrow {}^1\Sigma$ type, where the lower level was also the lower level of the visible $^1\Sigma \rightarrow {}^1\Sigma$ bands, and that their vibrational structure could be represented by the equation

$$\nu_o = 10542\cdot91 + 1127\cdot771\,(v'+\tfrac{1}{2}) - 8\cdot4007\,(v'+\tfrac{1}{2})^2 + 0\cdot03913\,(v'+\tfrac{1}{2})^3$$
$$- 1486\cdot87\,(v''+\tfrac{1}{2}) + 11\cdot70\,(v''+\tfrac{1}{2})^2 \qquad \text{......(3).}$$

Herzberg concluded that the ultra-violet and infra-red systems had the $^1\Pi$ level ($\omega_e = 1128$ cm.$^{-1}$) in common and that the higher value of the vibrational frequency given by the ultra-violet bands ($\omega_e = 1136$ cm.$^{-1}$) was due to the fact that equation (2) applied to heads, whilst equation (3) applied to origins. The deviation shown by the value obtained from the heads is to be expected, since the moment of inertia of the molecule in the two states is not very different (as is shown by the values of the vibrational frequencies of the two states) and the bands degrade to the red.

Equation (2) does not account for a number of heads commencing at λ 3364 and running to the red. These bands formed the strongest feature of the ultra-violet spectrum as we observed it, but no analysis was possible until the side sequences had been obtained. These are extremely faint, even on our heaviest plates, the (+1) sequence overlaps the (−1) sequence of Bengtsson's system, and the heads are very poorly defined. It was found possible, however, to fit these heads by the formula

$$\nu_h = 29753\cdot0 + 1083\cdot1\,(v'+\tfrac{1}{2}) - 11\cdot1\,(v'+\tfrac{1}{2})^2 - 1148\cdot7\,(v''+\tfrac{1}{2})$$
$$+ 11\cdot6\,(v''+\tfrac{1}{2})^2 \qquad \text{......(4).}$$

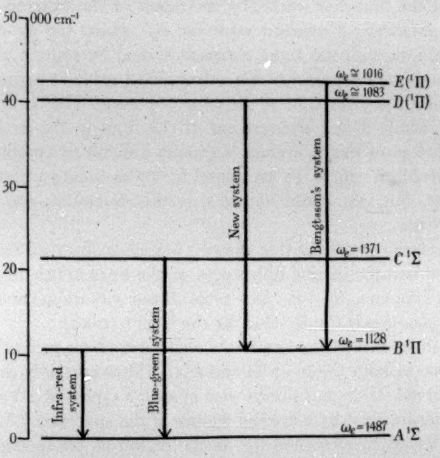
The fit given by this equation is shown in the table.

Table. Vibrational analysis of the D → B $^1\Pi$ system. (Wave numbers in brackets calculated from equation (4))

v' \ v''	0	1	2	3
0	29720·8 (29720·3)	28595·7 (28594·9)	—	—
1	30783·6 (30781·1)	29655·3 (29655·7)	28551·6 (28553·3)	—
2	—	30689·6 (30694·5)	29593·7 (29592·1)	28512·8 (28513·1)
3	—	—	30610·0 (30608·5)	29530·2 (29529·5)

It seems likely that the lower level here is the $^1\Pi$ level already discussed. The vibrational frequency (1148·7 cm.$^{-1}$) is higher again than the one obtained from the other ultra-violet system, but in the present case the moments of inertia in the two states are much more nearly equal (as is shown by the closer approach to equality of ω' and ω'') and therefore the heads will be formed at very considerable distances from their origins, and this distance ($\nu_h - \nu_0$) will differ markedly from sequence to sequence.

The electronic levels and transitions now known in the BeO spectrum are shown in the figure. In view of the unknown character of the two highest states, and in order to simplify future reference to the various systems, we have used the notation employed by Jevons ((5), pp. 76, 284) to describe the BeO levels but have used the labels D and E tentatively put forward by him in a rather different way.



§ 4. ROTATIONAL STRUCTURE

The nature of the two states D and E cannot be decided until a rotational analysis has been carried out in each case for bands involving these levels. We were hopeful that it would prove possible to do this for the new D → B $^1\Pi$ system since a considerable amount of structure was visible on our plates. Unfortunately, owing to the weakness of the side sequences, it was only found possible to study the main ($\Delta v = 0$) sequence. But even here unmistakable branches were only discernible in the (0, 0) band. This was due to the fact that, at the high temperature of either the arc or the spark, the maximum intensity of the various branches occurs at high values of J, so that the higher members of the $\Delta v = 0$ sequence were overlaid by branches from other bands to such an extent that analysis was impossible.

However, even with only the (0, 0) member of this sequence available, a definite conclusion as to the constants of the upper level involved in this system would have been possible had combination differences for the lower ($B\,^1\Pi$) level been obtainable from the analysis made by L. Herzberg of the $B\,^1\Pi \rightarrow A\,^1\Sigma$ system. Unfortunately, however, the analysis of this $B\,^1\Pi \rightarrow A\,^1\Sigma$ system gave only the B-values of the $B\,^1\Pi$ level from $v=2$ to $v=7$, although by extrapolation of the equation for these B-values B_0 should be 1·342. Similarly by extrapolation $D_0 = -7{\cdot}7 \times 10^{-6}$. Knowing these values we can calculate the term differences for the $B\,^1\Pi\,(v=0)$ state.

These same differences should, of course, have been obtainable from the (0, 0) band of the $D \rightarrow B\,^1\Pi$ system. We observed two branches in this (0, 0) band but could get no definite agreement between the differences obtained from these branches and the calculated differences. We did, however, succeed in getting an arrangement of the branches which, on treatment of the differences in the usual manner, gave practically a constant value for B_0', whilst the B_0''-values were too high (in comparison with the 1·342 obtained above) by some 2 per cent for low J-values although they approached the extrapolated value at higher J-values, and settled down within 0·5 per cent of the expected value. The constancy of the B_0'-values is inexplicable if our arrangement of the lines in the branches is faulty, whilst the behaviour of the B_0''-values is equally difficult to explain if our analysis is correct. It could, of course, be accounted for by assuming a perturbation of the $B\,^1\Pi\,(v=0)$ level, but this would not be a justifiable assumption in view of the paucity of our data.

We are, however, convinced that the two branches observed in the (0, 0) band are genuine, and that the second differences of the lines in the branches enable us to estimate the difference $B_0' - B_0''$ as $-0{\cdot}06$. If our vibrational analysis is correct, then $B_0' = 1{\cdot}28$, i.e. this is the B-value for the level $D\,(v=0)$.

As regards the nature of the two levels D and E it seems probable that both are of $^1\Pi$ type, since in both the $D \rightarrow B\,^1\Pi$ and $E \rightarrow B\,^1\Pi$ systems only single heads have so far been observed. If the transitions were of such a type that $\Delta\Lambda \neq 0$, then double heads would, presumably, be a marked feature of the spectrum. Actually the only head of any intensity unaccounted for, in the region of the spectrum that we are considering, is one at 29,683 cm^{-1}. It is not certain that this is a head. It may be an atomic line, but we have been unable to account for it as such. Its appearance is that of a moderately strong line diffuse to the red. It is possible that this is a Q head associated with our (0, 0) head, which is presumably of R type, at 29,720 cm^{-1}. According to our conclusion that the transition is probably of $^1\Pi \rightarrow {}^1\Pi$ type, and that the B_0-values are as suggested earlier, a weak Q head might be observable in this vicinity.

Finally, it is of interest to note that some very heavily exposed plates of the $C\,^1\Sigma \rightarrow A\,^1\Sigma$ system have been obtained and that no evidence for the existence of an isotope Be^8 is obtainable from them.

§ 5. ACKNOWLEDGMENTS

We take this opportunity of thanking Prof. Bragg for the facilities that have been placed at our disposal and Prof. Curtis for kindly reading the manuscript.

REFERENCES

(1) Bengtsson. *Ark. Mat. Astr. Fys.* A, **20**, 1 (1928).
(2) Rosenthal and Jenkins. *Phys. Rev.* **33**, 163 (1929).
(3) L. Herzberg. *Z. Phys.* **84**, 571 (1933).
(4) Jevons. *Proc. roy. Soc.* A, **122**, 211 (1929).
(5) Jevons. *Report on Band Spectra of Diatomic Molecules.*

535·33·03:546.17

THE EFFICIENCY OF EXCITATION OF THE NITROGEN FIRST POSITIVE BANDS BY ELECTRON IMPACT

By S. E. WILLIAMS, M.Sc., 1851 Exhibitioner (Sydney),
Wills Physical Laboratory, Bristol

Communicated by Prof. A. M. Tyndall, December 18, 1934. *Read in title March* 1, 1935.

ABSTRACT. A determination has been made of an upper limit to the effective target area of the nitrogen molecule for excitation to the triplet levels by electron impact. Observations were made on the visible groups of the first positive system by the use of an electron tube previously described. By calibrating the plate with a lamp, the absolute emission of which was found in comparison with a black-body radiator, the rate of emission of quanta in each group was calculated, and with the aid of the radiometric analyses available was extended to the whole of the first positive system. The cross-sectional area, for excitation by electrons with energy of about 14 volts, is, in terms of the area of the first Bohr orbit, found to be 5×10^{-2}, with an estimated probable error of 35 per cent.

In many problems connected with the electric discharge in nitrogen, the existence of metastable molecules in the $A\,^3\Sigma$ state plays an important role. It is therefore of interest to form some estimate of the rate of supply of these excited molecules by electron-collision under controlled conditions, i.e. to estimate the effective target area which a nitrogen molecule presents to an electron for the purpose of being excited to the metastable level.

The $A\,^3\Sigma$ level may be excited either directly or by cascade from the other triplet levels. The probability of direct excitation cannot be estimated spectroscopically because of the low intensity of the $A\,^3\Sigma \rightarrow X\,^1\Sigma$ intercombination system and also of the small chance of a metastable molecule making this transition before its energy is otherwise dissipated. On the other hand, cascade excitation is accompanied by emission of the first positive band system, and an estimate of the number of quanta emitted in this system gives at once the number of molecules in the $A\,^3\Sigma$ state produced in this way under the given experimental conditions.

This note gives an account of an absolute determination of the effective target area of the nitrogen molecule for production of the first positive system by collision with electrons of such energy that the emission of radiation in this band system is approximately a maximum; i.e. an upper limit is given to the effective cross-section for production of molecules in the $A\,^3\Sigma$ state by cascade. At the same time this result is also a rough estimate of the maximum efficiency of production of metastable molecules by both processes, because, although direct excitation might conceivably be much more important when electrons of smaller energy are present, since the $\mathrm{A}\,^3\Sigma$ excitation function has in all probability a maximum within a few volts of its excitation potential, the possibility of excitation by cascade is under these conditions correspondingly reduced. The relative orders of magnitude of the cross-sectional areas for excitation of molecular and atomic levels are also of interest.

The apparatus previously used to investigate the excitation function of the second positive bands has been described elsewhere(1). It consists of an indirectly heated cathode, slit system, and field-free observation chamber, through which nitrogen was allowed to pass in a continuous stream at a pressure of about 0·002 mm. The maximum electron energy was about 15·75 volts, and 90 per cent of the electron current was comprised in a range of about 4 volts below this value. An image of the electron beam was focused on the slit of a Hilger constant deviation spectrograph.

The radiation-intensity was small and, with the Ilford hypersensitive panchromatic plates used, exposures of about an hour were necessary. These plates are, of course, sensitive only to the visible of the several groups of bands in the first positive system. But Pfund(2) and Poetker(3) have published radiometric analyses of the complete system, and it was therefore found more convenient to deduce from them the proportion of the total energy of the system contained in each group than to use special infra-red plates and photograph the whole system. The plates were calibrated by means of continuous spectra of equal exposure-times and varying slit-widths, obtained from a lamp run at a known colour temperature and placed at a convenient distance from the spectrograph with no intervening optical system.

As the slit of the spectrograph was sufficiently wide to ensure complete mergence of the rotational structure, the images could be treated as a continuous spectrum. Owing to the effect of spreading discussed in another paper(1), the emission of light was not confined to the dimensions of the electron beam, so that the image received on the plate was a blurred patch with no sharply defined boundaries. This was photometered in a direction normal to the dispersion at intervals of 1 mm. measured parallel to the dispersion, the centre of one vibration band, just resolved, being taken as datum point. With the aid of the plate-calibration curves for each of these cross-sections (necessary owing to the rapid change in plate-sensitivity in this region) intensity contours could be built up, and the total intensity of the group could be found by integration. This was then expressed in equivalent intensity units per square millimetre; i.e., as if the whole of the intensity in the group were integrated and averaged over one square millimetre centred on the given wave-length, the value found being regarded as equivalent to that of the continuous spectrum with a slit of appropriate width.

For lack of absolute calibration of the tungsten lamp, a comparison was made, by means of a Hilger-Nutting photometer, with the cavity radiation from an alumina cylinder at a temperature of about 1000° C. Assuming that the lamp radiates in the manner of a black body, the absolute temperature of which is equal to the colour temperature of the lamp, it is possible to calculate for the latter an effective area of the emitting surface which, of course, varies with the wave-length owing to the changing spectral emission of tungsten. The appropriate value of this quantity is then used in Planck's radiation formula to calculate the rate of emission of quanta of a given wave-length, the spectral range $d\lambda$ in the formula being determined by the dispersion of the spectrograph at this point, since the length of spectrum considered on the plate is 1 mm.* The number of quanta falling on a square millimetre of plate then depends on the distance of the lamp from the slit and on the dimensions of the latter, i.e. its width per division and its length effective in producing a spectrum 1 mm. in depth, as given by the reciprocal of the magnification of the camera. Let this number be n. Then if the equivalent intensity of the band, in divisions of slit-width, is m, and the ratio of exposure times of band and continuous spectrum is R, the rate N at which quanta are photographed in the band is equal to nm/R^p, where p is the Schwarzchild factor for the plate, and is the weighted mean of all observations of intensity in the band, owing to its variation with plate-blackening. From the geometry of the optical system used in photographing the beam the total number of quanta emitted in the band can be determined, and an analysis of the radiometer curves enables a calculation of a similar quantity for the whole system.

The measured gas pressure and estimated temperature give the number of molecules per cubic centimetre, and the electron current is known. From a knowledge of the slit-width and relative positions of the slit, focusing lens and beam, the length of the latter under observation can be found. The calculation of the effective cross-section is then quite simple.

The table shows the principal data used in four calculations of the effective cross-section. They are taken from each of the two visible groups of first positive bands, obtained in photographs with exposures of 1 and 2 hours respectively. The effective area is given in terms of πa_0^2, the area of the first Bohr orbit, a_0, having the value 0·528 A.

Group	Equivalent intensity (slit-div./mm²)	Percentage of total quanta in system	Exposure (min.)	Pressure (mm. Hg. $\times 10^{-3}$)	Current (mA.)	$\pi r^2/\pi a_0^2$
6500 A.	29·3	19	120	1·55	0·98	$5{\cdot}26 \times 10^{-2}$
	16·5	19	60	1·7	1·0	$5{\cdot}14 \times 10^{-2}$
5900 A.	16·7	4·5	120	1·55	0·98	$5{\cdot}17 \times 10^{-2}$
	8·3	4·5	60	1·7	1·0	$4{\cdot}85 \times 10^{-2}$

Mean of $\pi r^2/\pi a_0^2$, $5{\cdot}1 \times 10^{-2}$.

The accuracy of this result is difficult to assess, owing to the large number of independent factors which enter into the problem. Whilst the consistency of the four determinations is very satisfactory, the extreme deviation being only 5 per cent, it is thought that the probable error should be placed at the higher figure of 35 per cent. Of this, 8 per cent is accounted for in the determination of the absolute emission from the lamp, 10 per cent in the equivalent intensities, and 10 per cent in the relative numbers of quanta in each group of the system.

---

* Since comparisons with the black body were made at the wave-lengths required in the calculation a knowledge of the colour temperature of the lamp is rendered unnecessary, as the exponential factor involving this quantity cancels out.

I am indebted to Dr C. H. Johnson, of the Chemical Department, for assistance in the use of a Hilger-Nutting spectrophotometer, and to the Director of the Ilford Research Laboratories for a determination of the Schwarzchild factor of the photographic plates. In particular, I desire to express thanks to Dr E. T. S. Appleyard for much helpful advice, discussion of the problem, and assistance in checking the calculations.

REFERENCES

(1) THOMPSON and WILLIAMS. *Proc. roy. Soc.* A, **147**, 583 (1934).
(2) PFUND. *J. opt. Soc. Amer.* **9**, 193 (1924).
(3) POETKER. *Phys. Rev.* **30**, 812 (1927).

621.327.31 535.337:546.49

ABSOLUTE INTENSITIES IN THE SPECTRUM OF QUARTZ MERCURY ARCS AND THEIR VARIATION WITH TEMPERATURE-CHANGES OF THE SURROUNDING AIR

By A. J. MADDOCK, M.Sc., A.Inst.P.

Received November 1, 1934. *Read in title March* 15, 1935.

ABSTRACT. Variations in intensity of the more important lines in the mercury arc spectrum are investigated as the temperature of the air surrounding the burner is increased. The intensity and arc watts reach maxima but not quite together. Empirical relations are found connecting intensity and electrical input. Absolute intensities of some thirty lines between 7000 and 2300 A. emitted by commercial quartz mercury-vapour burners are tabulated and some theoretical aspects of the results are discussed.

§ 1. INTRODUCTION

Mercury-vapour burners are usually operated in some form of housing, and as the intensity of radiation emitted is susceptible to changes of temperature it is of interest to know how it varies as the temperature of the surrounding air increases. In different housings a given burner will attain equilibrium at different electrical ratings depending on the ventilation.

In the investigation to be described, the temperature of the air around several quartz mercury-vapour burners was varied and the intensity of some of the most prominent lines in the spectrum was determined, the supply voltage and ballast resistance being maintained constant. Results have been quoted bearing upon the variation in intensity when the supply volts are varied(1), and Harrison and Forbes(2) have described a special type of lamp wherein all the variables were separable and the effect of varying either the voltage across or the current through the tube was determined. Variations of intensity with change of voltage-gradient, for constant current, have been given by Taylor(3). Our object was to investigate the changes that might occur in the practical operation of commercial mercury-vapour burners, i.e. the intensity-variations resulting solely from temperature-changes such as are due to operation of the burners in housings differing in ventilation.

A complete investigation was also made of the absolute intensities of some 30 lines between 7000 A. and 2300 A. in the spectrum from quartz mercury arcs, wherein the intensity due to the whole of the burner at a distance of 1 metre was determined. From such results therapeutic, bactericidal and other effects can be deduced if desired. Some theoretical aspects of the results are discussed.

§ 2. PROCEDURE

The burner was placed 117 cm. from the entrance slit of a spectrometer so that radiation from the whole arc-length entered the instrument; a thermopile, after the exit slit, connected to a Paschen galvanometer (the combination having been calibrated by means of a carbon filament lamp standardized by the National Physical Laboratory) gave the intensity after transmission through the spectrometer*; losses in this latter were determined experimentally by means of an auxiliary monochromator and thus the actual intensity at the entrance slit could be determined for any wave-length.

The temperature of the air surrounding the burner was decreased by means of a fan or increased by means of an electric heater immediately below the burner. It was not practicable nor was it considered necessary to determine the actual temperature of the burner at any point; as will be seen later, the conditions of operation can be inferred from the electrical characteristics.

The exit slit was set wider than the collimator slit in order that the intensity of lines close together might be integrated by the thermopile and the intensity for the group obtained, rather than that for each individual line comprising the group. The wave-lengths studied were 5790, 5770 A. together; 5461 A.; the 4358-A. group; the 3650-A. group; the 3126-A. group; 3022 A.; the 2652-A. group; and 2535, 2537 A. together. Their intensities were determined for various equilibrium conditions of the burner brought about by cooling or heating as described, and the electrical conditions also were noted.

§ 3. RESULTS AND DISCUSSION

Variation of intensity. As is generally known, the intensity of all the spectral lines increases as the ambient air-temperature increases, and at the same time the arc watts increase, eventually passing through a maximum and then decreasing; the intensity, however, continues to increase after this maximum of power has been reached until, on further heating of the surrounding air, this also soon reaches a maximum and then begins to decrease. Typical curves showing the variation of intensity with arc watts due to changes in air-temperature for a 2·5-A. burner† (operated on 240-V. mains with a series resistance of 36 Ω.) are given in figure 1, and the maxima for watts and intensity are clearly shown. The primary variable is, of course, the temperature of the air in the immediate vicinity of the burner, and related variables are the temperature of and pressure within the burner itself.

The arc voltage and current are good indications of the conditions under which

---

* A null method was used as being more practicable in overcoming disturbances and drift. A series and shunt resistance circuit injected an opposing current into the galvanometer circuit, switching taking place automatically as the light-shutter was raised; the balancing potential across the main potentiometer, as read on a sub-standard voltmeter, gave the required experimental observations proportional to the intensity.

† Normal operating conditions for this burner, with the specified mains voltage and ballast resistance, is 2·5 A. which is the current-value for ventilation given in a normal housing with free air circulation.

the burner is operating, for as the temperature of the air is raised from the coolest to the hottest degree practicable, the arc volts increase whilst the current decreases.
E_s, R Since the supply volts, E_s, and ballast resistance, R, are maintained constant,
E, i the arc volts E and current i are related linearly, the characteristic having a negative slope numerically equal to the series resistance R. For

$$E=E_s-Ri \qquad \ldots\ldots(1).$$

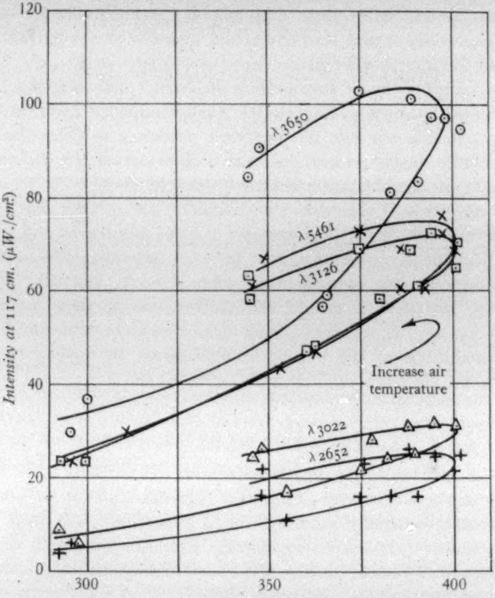


Figure 1. Arc watts and intensity due to temperature-variation of air around burner no. 844. Mains voltage, 240 V.; ballast resistance, 36 Ω.

P Now the arc watts $P=Ei=E_si-Ri^2$; differentiation shows that P is a maximum when $i=E_s/2R$ giving

$$P_{\text{max.}}=E_s{}^2/4R \qquad \ldots\ldots(2).$$

$P_{\text{max.}}$ as found by experiment was 400 W. and is also 400 by substitution in equation (2) of $E_s=240$ V. and $R=36\ \Omega$.

The maximum intensity occurs at a temperature such that the arc watts are a little less than their maximum; thus maxima for λλ 5780, 5461, 4358 occur at about 385 W. whilst maxima for λλ 3126, 3022, 2652, 2536 occur at about 396 W. and a maximum for λ 3650 occurs at about 380 W. In the ultra-violet region of wave-lengths from λ 3126 downwards, the maxima of intensity all occur very near the maximum wattage, the peak being sharper than for long wave-lengths. For maximum intensity, therefore, this burner should be operated with such ventilation that the arc watts lie between 385 and 400 at the higher voltage, i.e. between 120 and 140 V. with corresponding currents of 3·3 and 2·75 A.; the normal operation is at 2·5 A., which occurs at 150 V. (375 W.), and though the intensities are not quite at their maxima they are nearly so and operation at any voltage between 120 and 150 will be satisfactory and ensure that only small changes of intensity shall occur with any temperature-fluctuations. Under these conditions the intensity is only subject to small variations for quite large temperature-changes, and this range is the most suitable for operation. That all lines exhibit decreased intensities when such a burner is operated under extra hot conditions is shown by comparison of columns 2 and 3 in table 2.

Exactly similar results were obtained with a 3·5-A. burner and 28-Ω. series resistance, the maxima of arc wattage again not producing maxima of intensity. Maximum power was 515 W. by experiment and 514 by substitution in equation (2). The maximum intensity for $\lambda\lambda$ 5461, 4358 and 3126 occurred at about 505 W., that for $\lambda\lambda$ 3650, 3022, and 2536 at 510 W.; thus, in this case also, for maximum intensity the burner should be operated at such a temperature that the voltage lies between 120 and 140 with corresponding current-values of 4·28 and 3·56 A., normal operation being at 3·5 A., 141 V. and 494 W. These two types of burner have the same arc-length, about 12 cm., and it is to be noted that maximum intensity occurs for both types within the same voltage-range, i.e. with the same value (10 to 11·7) of volts per centimetre, the greater arc current being responsible for the greater actual intensity.

Between arc voltages of 80 and 140 the foregoing experimental results can be expressed by the relation

$$I = CiE^n \qquad \ldots\ldots(3),$$

where I is the intensity of radiation of some particular wave-length and C and n are constants corresponding to that wave-length but differing for each spectral line. This is a relation between I, i and E when the supply volts and series resistance are constant and the only variable is the temperature of the air around the burner, i and E being further related by equation (1). Plotting of $\log_{10}(I/i)$ against $\log_{10} E$ results in the type of curve shown in figure 2, where the values for some of the wave-lengths of the 2·5-A. burner are given, these being typical of all the results. The linear relation is quite well defined between 80 and 140 V. but below this the curve departs considerably from linearity. Above 140 V. there is a marked change in slope, but the exact form is difficult to determine since it was not found possible to extend the observations much above the points given. Further rise of temperature increased the pressure in the arc space so much that the arc was extinguished.

I, C, n

Values for n and C as determined from the linear portion of the curves are given in table 1, I being in μW./cm.^{2} at 117 cm. distance and i and E in amperes and volts respectively.

Table 1. Values of *n* and *C* for equation (3)

λ	2·5-A. burner		3·5-A. burner	
	n	*C*	*n*	*C*
5780	1·93	0·0017	—	—
5461	1·74	0·0049	1·66	0·008
4358	1·75	0·0040	1·57	0·0115
3650	1·99	0·0021	1·86	0·0046
3126	1·77	0·0045	1·42	0·026
3022	1·92	$0{\cdot}0008_8$	2·11	$0{\cdot}0004_6$
2652	2·24	$0{\cdot}0001_4$	1·76	$0{\cdot}0016_5$
2536	1·88	$0{\cdot}0011_5$	1·48	0·0092

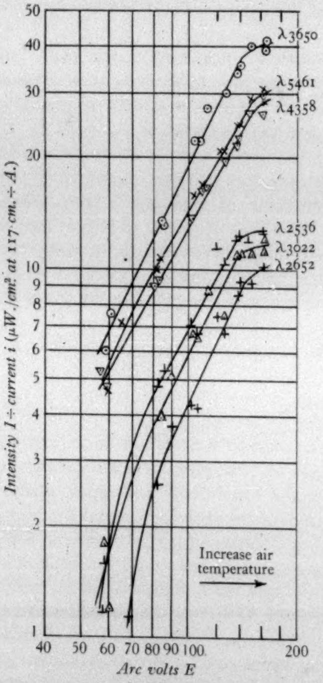


Figure 2.

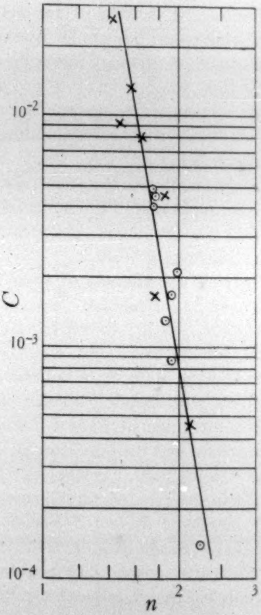


Figure 3.

Figure 2. Relation between intensity/current and arc volts as temperature of air around burner is increased. Burner, mains voltage and resistance as in figure 1.

Figure 3. Relation between *n* and *C*. × Burner no. 160, 3·5-A. type; ⊙ burner no. 844, 2·5-A. type.

It will be seen that n is different for each wave-length, and this shows that the relative distribution of intensity throughout the spectrum will vary with the operating conditions; thus, for example, as λ 3650 has a large slope or large n the excess of its intensity over that of other wave-lengths will be greater at high temperatures than at low; a similar observation holds for λλ 5780, 3022 and 2652. C and n appear to be related, for on plotting $\log_{10} C$ against $\log_{10} n$, figure 3, we find that the points are distributed more or less evenly about a straight line which apparently fits the data from both the 2·5-A. and the 3·5-A. burners owing, no doubt, to the equal arc-lengths. From figure 3 we have

$$C = 1{\cdot}8 n^{-0{\cdot}0082} \qquad \ldots\ldots(4).$$

Differentiation of equation (3) gives $I_{\text{max.}}$ as occurring at about 150 V. which, as we have seen, is beyond the region for which the relation holds, and hence this operation is inapplicable.

A somewhat similar expression is obtained by Asada⁽⁴⁾ for the variation of intensity with change of terminal voltage, though full details of his results are not given. He obtains $I = Ki\,(E - E_0)^n$ subject to the condition that the potential difference be not very small, and states that the radiations of longer wave-lengths increase more rapidly than those of shorter. E_0 is the arc voltage on first striking.

E_0

Absolute intensities. The results presented in the preceding section were obtained as part of a more complete investigation wherein were obtained the absolute intensities of all the main wave-lengths between 7000 and 2300 A. emitted by several mercury vapour burners and measurable by the radiometric method already described. Most of the intensities were determined for a 5-cm. length of arc at 45 cm. from the collimator slit, and by means of an open thermopile and three filters (passing λ 5461, λ 4358 and all wave-lengths less than 1·4μ) a mean ratio was determined for obtaining the intensity due to the whole arc length at 100 cm. distance. Data thus obtained are summarized in table 2. Columns 3, 4, 5 relate to burners operated under such ventilation conditions as to permit operation with their normal current. Column 2 for no. 793 gives the intensity for this burner operated under much hotter conditions, such as obtain in a closed lamp-house with little air-circulation, and the decrease that results is at once evident. No. 844 was a new burner whilst no. 793 had been in operation before and the loss of intensity for the shorter wave-lengths was becoming apparent.

Since λ 2537 spreads out on the long-wave side to about λ 2587 and the slit was insufficiently wide to accommodate this range, a factor was obtained from microphotometer records of this region whereby the total intensity of all of λλ 2537 and 2535 could be obtained from measurements made with the spectrometer set at λ 2536. From such records, also, the amount to be deducted from the reading at λ 2576 due to λ 2537 was deduced, the intensity-values in the table being those of λ 2576 alone.

These intensity-measurements cover a wide range of wave-lengths and it is interesting to note the decay of intensity in the sharp ($2^3P_{0,1,2} - m^3S_1$) and diffuse ($2^3P_{0,1,2} - m^3D_{1,2,3}$) series. In figure 4 are plotted the logarithm of the intensity

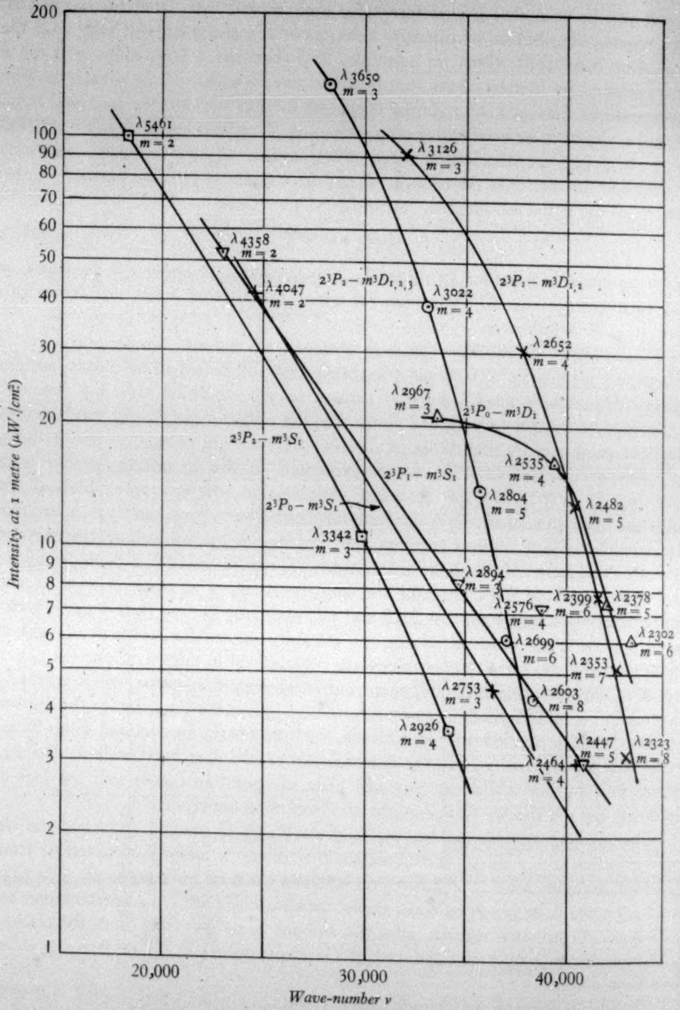


Figure 4. Relation between intensity and wave-number for the sharp and diffuse series of the mercury spectrum. Burner, mains voltage, and ballast resistance as in figure 1; arc voltage 146·4, arc amperage 2·49.

against wave-number for the lines belonging to these series; these are for burner no. 844 and are typical of all the results. Smooth curves can be drawn through practically all the points in a given series and this indicates that experimental errors were small. These results are in agreement with those of McAlister⁽⁵⁾, and the curves, like his, appear to approach asymptotic lines of approximately the same slope for the two sets of series; this slope is greater for the diffuse series, a fact which indicates a lower temperature of the excited atoms for the 2^3D than for the 2^3S levels as was pointed out by Hulburt⁽⁶⁾. Hulburt argues that if there is a sort of temperature-equilibrium in the mercury arc, then the intensity I of radiation of wave-number ν is

$$I = \alpha \exp\left(-hc\nu/kT\right) \qquad \text{......(5)},$$

where α represents the various probabilities of state and of transition and the usual meanings attach to the other symbols. If α is constant for the lines of a series, the logarithm of the experimental intensity plotted against wave number should be a straight line for the series. Our work and that of McAlister shows that α is not constant but varies with ν, decreasing as ν increases. The change of α with ν is greater in the diffuse series and small in the sharp series, the lines of the latter in some cases being linear, although we have insufficient points to be able to argue this rigorously. From the high-frequency ends of the curves, where they have become practically linear, we deduce approximate temperatures of 6700° K. for the sharp and 2500° K. for the diffuse series, these estimates being based on the four sets of intensity-measurements given in table 2.

In the plot of log I against ν, the intensities of λ 4358 and λ 4339; λ 4078 and λ 4047; λ 2537 and λ 2535 were separated as these pairs are not of the same series. This was done either by means of the microphotometer records or by taking intensity-readings at the separate wave-lengths and calculating the amount to be deducted in respect of the overlap of the neighbouring line. λ 2576 does not fit the general curve in any of the four sets of results, though the linearity of the series to which this belongs is borne out by the fit of λ 2447.

§ 4. ACKNOWLEDGMENTS

Thanks are due to the Thermal Syndicate Ltd. for permission to publish this paper. The investigation was carried out in their Laboratories and the arcs experimented upon were their standard 2·5-A. and 3·5-A., 220–250-V., d.-c.-type quartz mercury vapour burners.

REFERENCES

(1) Coblentz, W. W., Long, M. B. and Kahler, H. *Sci. Pap. U.S. Bur. Stand.* No. 330 (1918).
(2) Harrison, G. R. and Forbes, G. S. *J. opt. Soc. Amer.* **10**, 1 (1925).
(3) Taylor, P. B. *J. Franklin Inst.* **207**, 95 (1929).
(4) Asada, T. *Nature*, Lond., **127**, 254 (1931).
(5) McAlister, E. D. *Smithson. misc. Coll.* **87**, No. 17, Publ. 3187 (1933).
(6) Hulburt, E. O. *Phys. Rev.* **32**, 593 (1928).

Table 2. Radiant energy flux (μW./cm.^{2}) at 1 metre from the centre of quartz mercury vapour burners

λ	2·5-A. type. 240-V. mains. 36-Ω. series resistance			3·5-A. type. 240-V. mains. 28-Ω.
	No. 793		No. 844	No. 160
	Extra hot 162·0 V. 2·03 A.	Normal 147·3 V. 2·49 A.	Normal 146·4 V. 2·49 A.	Normal 141·5 V. 3·49 A.
7082 6907	—	—	5·7	9·7
6234 6149 6124 6073	—	—	2·2	5·3
5790 5770	68·7	88·4	95·0	176·0
5461	73·4	91·0	98·9	161·0
4916	2·2	2·0	2·2	3·5
4358 4339	67·9	83·6	83·7	135·0
4078 4047 3984	33·1	41·3	46·0	72·0
3906	—	0·9	1·3	4·7
3704 3663 3655 3650	90·3	116·0	134·0	219·5
3342	5·9	8·6	10·8	17·7
3132 3126	56·7	67·4	91·1	140·3
3022	18·2	25·1	39·2	69·0
2967	11·5	13·6	21·1	32·7
2926	—	1·2	3·6	6·8
2894	3·6	4·9	8·3	11·4
2804	5·2	6·5	14·1	21·2
2753	—	1·4	4·6	7·2
2699	2·5	3·6	6·1	13·5
2655 2654 2652 2640	12·0	16·8	30·6	53·3
2603	—	—	4·4	7·5
2576	2·5	2·6	7·2	14·0
2537 2535	18·1	24·0	50·9	76·4
2482	4·1	6·0	12·9	20·5
2464	—	—	3·1	—
2447	—	—	3·1	—
2399	2·0	2·9	7·7	11·8
2378	1·9	3·0	7·5	8·4
2353	—	—	5·2	6·4
2323	—	—	3·2	—
2302	—	—	6·1	—
Total	480	611	809	1295
Total for λ 3132 and shorter wave-lengths	138	179	330	500

535.338.4 : 546.881

THE BAND SPECTRUM OF VANADIUM OXIDE

By P. C. MAHANTI, M.Sc.

Communicated by Prof. P. N. Ghosh, January 12, 1935. *Read in title March* 15, 1935.

ABSTRACT. The spectrum of VO has been photographed under high dispersion. Experimental evidence confirming that the emitting molecule is VO has been secured. The vibrational analysis of the band system has been extended to $v' = 8$, and approximate values of the vibrational constants have been calculated from the new band-head data. The rotational structure-analysis of the bands (0, 1), (0, 0) and (1, 0) shows the presence of two R and two P branches with a short Q branch. The rotational constants of the molecule have been evaluated. It is suggested that the band system is due to a $^2\Delta \rightarrow {}^2\Delta$ transition.

§ 1. INTRODUCTION

The spectrum of vanadium oxide, VO, consists of well-marked red-degrading bands which extend from violet to far red. As early as 1895 Demarçay(1) first observed them in the spark through a solution of vanadium fluoride. But the wave-length data of only a few of their heads were recorded by the early investigators(2). Using low dispersion, Ferguson(3) has recently measured a large number of these bands.

The vibrational quantum-analysis of these bands was first undertaken by Mecke(4), who had identified only four heads, viz. (1, 0), (0, 0), (0, 1) and (0, 2) from the old data. Ferguson has included as many as thirty-one bands in the system. This enabled him to calculate approximate values of the vibrational constants of the molecule. He has published, however, no photograph of his spectrum.

The main object of the present investigation is to make a rotational structure-analysis of the bands. Of the band spectra known so far to be associated with the oxides of the elements of the transition groups, a structure-analysis has been made only in the case of the blue-green bands of titanium oxide and it might be interesting to make such a study for those of vanadium oxide. For this purpose, first- and second-order spectrograms taken with a 21-ft. concave grating have been utilized. In addition to this, photographs taken with a Hilger E. 1 prism spectrograph, of the Littrow type, with interchangeable quartz and glass optical systems, revealed the presence of a number of bands which have not been recorded previously. The new data of wave-lengths obtained for the band heads have been used to calculate the values of the vibrational constants, which do not differ appreciably from those of Ferguson.

§ 2. THE IDENTIFICATION OF THE EMITTER

The fact that the same band spectrum is emitted, whether vanadium or its various salts are used in the core of the carbon arc in air, leads one to attribute it to the oxide of the metal but does not exclude the possibility of its being due to the hydride VH or the metal molecule V_2 itself. The rotational structure-analysis of the bands definitely rules out the hydride as the probable emitter, but some doubt is left as to whether V_2 might not be the molecule concerned. It was, therefore, considered desirable to obtain further experimental evidence regarding the identity of the emitter of the spectrum. Accordingly a water-cooled enclosed arc was used. The electrodes were of carbon, the lower one being cored to receive the metal or its salt. The arc could be satisfactorily maintained in different gases under varying pressures. It was found that the spectrum was emitted only in an atmosphere of either oxygen or air, while it was completely quenched when hydrogen or nitrogen was introduced into the arc chamber. This shows that the presence of oxygen is necessary to produce the bands and it may therefore be concluded that vanadium oxide, VO, is the emitter of the band spectrum in question.

§ 3. EXPERIMENTAL

The flame surrounding a carbon arc whose lower electrode contained metallic vanadium or vanadium salts was used as the source of radiation. Various salts such as vanadium chloride, vanadium phosphate, ammonium vanadate and vanadic acid have been used. The arc was operated at a current-density of 4 to 5 A. from 220-V. d.-c. mains.

For a preliminary survey spectrograms were obtained with the help of a Hilger E. 1 prism spectrograph. Ilford special rapid panchromatic plates were used. With an exposure of about 20 to 30 minutes, the bands were very well developed on the plates in the region extending from λ4400 to about λ7100. For photographing the bands beyond λ7100 dicyanine and kryptocyanine plates were used. The time of exposure was about 2 hours. The bands were then found to extend as far as λ8650. In the long wave-length region they were also photographed in the first order of a 15-ft. concave grating having 15,000 lines per inch.

For measuring structure lines of the bands, photographs were taken in the first and second orders of a 21-ft. concave grating set up in Paschen mounting and having 30,000 lines per inch with a 6-inch ruled surface. In the first order, an exposure of about 3 hours was sufficient to get the lines well developed for measurements. But in the second order the time of exposure was as long as 8 hours. Even then the lines, although just intense enough for measurement, were not sufficiently strong for reproduction.

Measurements of band heads and structure lines were made on several of the plates taken with the prism and grating spectrographs, figure 1. The dispersions

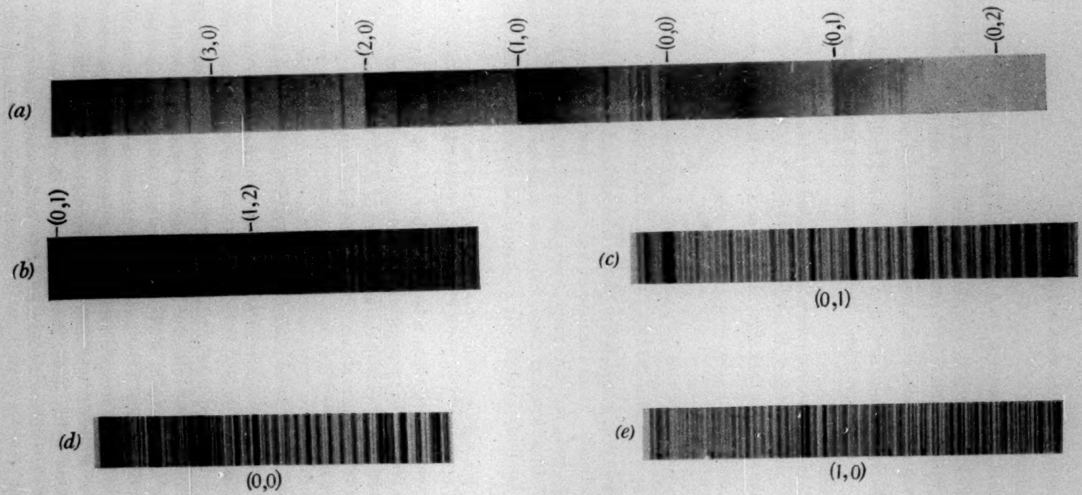

(a)
(3,0)
(2,0)
(1,0)
(0,0)
(0,1)
(0,2)
(b)
(0,1)
(1,2)
(c)
(0,1)
(d)
(0,0)
(e)
(1,0)

Figure 1. VO bands: (*a*) Hilger E. 1 quartz spectrograph; (*b*) 15-ft. grating, first order; (*c*, *d*, *e*) 21-ft. grating, first order.

given by these instruments were as follows:

Prism spectrograph, E. 1

with quartz optical system, 16 A./mm. at λ4400 to 55 A./mm. at λ7100,

with glass optical system, 6 A./mm. at λ4400 to 36 A./mm. at λ7600;

15-ft. grating spectrograph, 3·55 A./mm. in the first order;

21-ft. grating spectrograph, 0·62 A./mm. in the second order.

For standards of comparison, iron arc, neon and hydrogen lines were used for different regions of the spectrum. Measurements were made in the usual way with a Gaertner comparator.

Particulars of the observed bands are included in table 1, which indicates wave-lengths in air, wave-numbers *in vacuo* and relative intensities of the band heads. The data for the structure lines of (0, 1), (0, 0) and (1, 0) bands are given respectively in tables 2, 3 and 4.

§ 4. VIBRATIONAL ANALYSIS

As has already been stated, it was Mecke who initiated the vibrational quantum-analysis of the bands. From the data of Eder and Valenta, he assigned the heads at λλ5470, 5737, 6087 and 6478 respectively to (1, 0), (0, 0), (0, 1) and (0, 2) bands of the system. In addition to these, Ferguson included as many as twenty-seven bands, which are fairly intense. He evaluated the following vibrational constants in cm^{-1}:

$$\nu_e = 17498{\cdot}8,\quad \omega_e' = 864{\cdot}8,\quad \omega_e'x_e' = 5{\cdot}70,\quad \omega_e'' = 1012{\cdot}3,\quad \text{and}\quad \omega_e''x_e'' = 5{\cdot}27.$$

In the present investigation, a large number of new heads have been incorporated in the system. For comparison, Ferguson's wave-number data have been included in table 1 together with the assignment of v', v'' values to each band head.

Assuming that the heads are very close to the band origins, the equation derived for the R heads is

$$\nu = 17501{\cdot}3 + [863{\cdot}5\,(v'+\tfrac{1}{2}) - 5{\cdot}4\,(v'+\tfrac{1}{2})^2] - [1012{\cdot}7\,(v''+\tfrac{1}{2}) - 4{\cdot}9\,(v''+\tfrac{1}{2})^2].$$

The distribution of relative intensities of the bands in the system follows a typical wide Condon parabola.

It may here be noted that in addition to the bands included in the present system, Ferguson measured forty-five heads which he thinks may also be due to VO. They lie in the far red region of the spectrum and probably extend still towards the higher wave-length side. These bands could neither be fitted into an independent system nor incorporated into the existing one. Fourteen of them were, however, formed into four sequences and their first differences indicate that they might be related to the band system under consideration. In the present investigation these bands have not been obtained owing probably to the want of sensitiveness of the plates used.

Table 1

v', v''	λ in air and intensity	ν *in vacuo*		v', v''	λ in air and intensity	ν *in vacuo*	
		Author	Ferguson			Author	Ferguson
6, 0	4466·4 (0)	22383·1	—	2, 2	5837·3 (4)	17126·5	17125
7, 1	4510·1 (1)	22166·3	—	4, 4	5942·5 (1)	16823·3	—
8, 2	4554·1 (1)	21952·1	—	5, 5	5997·6 (1)	16668·7	—
5, 0	4632·7 (2)	21579·7	—	0, 1	6086·4 (8)	16425·4	16428
6, 1	4676·8 (2)	21376·2	—	1, 2	6139·1 (3)	16284·5	—
7, 2	4721·2 (2)	21175·1	—	2, 3	6191·6 (2)	16146·5	—
8, 3	4767·5 (1)	20969·1	—	3, 4	6247·2 (1)	16002·7	—
4, 0	4813·5 (3)	20769·1	20773	4, 5	6303·3 (1)	15860·3	—
5, 1	4858·5 (4)	20576·7	20578	5, 6	6360·4 (3)	15717·9	—
6, 2	4904·2 (4)	20385·0	20383	6, 7	6418·5 (1)	15575·7	—
7, 3	4951·2 (2)	20191·5	20191	0, 2	6477·8 (6)	15433·1	15433
8, 4	4999·1 (2)	19998·0	—	1, 3	6532·8 (6)	15303·2	15306
3, 0	5010·5 (4)	19952·5	19950	2, 4	6588·9 (3)	15172·9	15174
4, 1	5057·3 (4)	19767·9	19768	3, 5	6646·4 (2)	15041·6	—
5, 2	5104·1 (3)	19586·7	19586	4, 6	6706·0 (1)	14907·9	—
6, 3	5152·5 (3)	19402·7	—	5, 7	6766·2 (1)	14775·3	—
7, 4	5201·4 (2)	19220·3	—	6, 8	6827·6 (1)	14642·4	—
2, 0	5228·2 (6)	19121·7	19121	0, 3	6919·0 (4)	14449·0	14454
3, 1	5275·8 (6)	18949·2	18951	1, 4	6976·2 (4)	14330·5	14334
4, 2	5324·5 (6)	18775·9	18779	2, 5	7035·2 (3)	14210·3	14212
5, 3	5373·3 (2)	18605·4	—	3, 6	7095·6 (3)	14089·4	14098
6, 4	5425·1 (1)	18427·7	—	4, 7	—	—	13984
1, 0	5469·3 (9)	18278·8	18279	0, 4	7418·2 (3)	13476·6	13479
2, 1	5517·3 (5)	18119·8	18123	1, 5	7477·0 (2)	13370·7	13376
3, 2	5567·7 (4)	17955·8	—	2, 6	7538·2 (1)	13262·1	13270
4, 3	5618·0 (3)	17795·0	—	3, 7	—	—	13166
5, 4	5669·9 (2)	17632·1	—	0, 5	7986·0 (2)	12518·5	12522
6, 5	5723·8 (1)	17466·1	—	1, 6	—	—	12426
0, 0	5736·7 (10)	17426·9	17425	0, 6	8643·4 (2)	11566·3	11571
1, 1	5786·4 (1)	17277·1	—	1, 7	—	—	11489

§ 5. ROTATIONAL ANALYSIS

The strongest bands in the system, namely (1, 0), (0, 0) and (0, 1), were chosen for a study of rotational structure. Under high dispersion they show two heads very close to one another, although the resolution is incomplete in their neighbourhood. As one proceeds a little farther from the second head, two series of lines can be clearly followed. Each member of these series in its turn breaks up into two components giving rise to an extended set of four series of lines. *A priori* this leads one to think that the four series are probably due to R_1, R_2, P_1 and P_2 branches, although the individual series could not be assigned to any definite branch. As a first step towards the identification of the different series it was thought necessary to form all possible combinations between the members of any two series for each of the bands under consideration. Different sets of frequency-differences were thus obtained. With a view to selecting from them the proper combination differences between the lines of an R and a P branch, we next proceeded to ascertain probable values of $\Delta_2 T$. We had thus to assume first a value of B_0 which involves a knowledge of r_0 for the vanadium-oxide molecule. From Morse's relation, $\omega_0 r_0^3 \doteqdot$ a constant, one can calculate the value of r_0 for a molecule whose ω_0 is known, provided the magnitude of the constant can be properly chosen. The value of this constant was secured from the data of Christy$^{(5)}$ for the blue-green bands of titanium oxide. r_0 being known, B_0 was calculated and finally the expected values of $\Delta_2 T$ were obtained. With these values in mind, the frequency differences obtained from any two series of the (0, 0) band were compared with those from any two series of the (0, 1) band. When two sets of approximately equal values of $\Delta_2 T$, one from each band, were obtained, then the pair of series of the (0, 0) band were shifted by two units with respect to each other and were compared with an arbitrary pair of series of the (1, 0) band. It was found, however, that no combination relations could be satisfied except for a few members of the series. The above procedure was repeated for several other values of B_0 and thus of $\Delta_2 T$ in the neighbourhood of the one already tried. But each of these trials proved futile. This failure indicated that the values of B_0 tried were not the true values, and consequently the true values of $\Delta_2 T$ for the band lines could not be obtained. This supposition seems to be justified from the fact that, as Christy has remarked, the value of r_0 for the titanium-oxide molecule is decidedly small and is much less than the minimum value predicted by Birge. One may therefore expect that the value of $\omega_0 r_0^3$ is probably higher than that on which the above approximate values of B_0 are based. Hence the next procedure consisted in finding out whether for any sets of frequency-differences, irrespective of the magnitude of B_0, the combination relation could be satisfied between the three bands for any two of the four series of lines. After repeated trials pairs of such series were found for each of the bands, and in each pair the identification of R and P series was consequently ascertained and so also their J numbering. This helped us in identifying the two components of R and P branches. But it was difficult to decide which of them is to be associated with R_1, R_2 and P_1, P_2 series. It was then found that of the two series of lines which emanate a little distance away from the second head, one is a component of the R branch while the other is formed from the superposition of the second component of the R branch and of one component of the P branch. When each of these series is extrapolated towards the head, only one of them is found to form the first head while the other coalesces into it before reaching the second head. This shows evidently that the second head is not due to an R branch, and its head-like appearance may be attributed to the condensation of lines of a Q branch. It may further be noted that in band systems due to non-hydride molecules, R_1 and R_2 heads are separated only in cases where either the heads are formed by R-branch lines having high values of K or where the spin doubling is large. None of these criteria is fulfilled by the present analysis. On the other hand a further close inspection of the plates did not reveal the existence of a definite series which may be assigned to a Q branch. It is, therefore, probable that the Q branch is very short although it begins with appreciable intensity. It should be noted, however, that the superposition of the lines of the returning R branch also adds to the apparent intensity of the second head. From the values of B_v' and B_v'' evaluated from the analysis of the bands, the wave-numbers of the first few lines of a Q branch were calculated and they were found to lie very close to one another. It is quite probable that they would present a head-like appearance under the dispersion used. Besides this, the intervals between the first and second heads of (0, 1), (0, 0) and (1, 0) bands also satisfy the criterion for the identification of Q heads as they are respectively equal to 2·83, 2·64 and 2·41 cm^{-1}.

In tables 2, 3 and 4, wave-number data of the structure lines with their K numbering are given. They also include the mean values of the rotational term-differences for the upper and lower states, defined as follows.

In the upper state,

$$\Delta_2 T_1'(J) = R_1(J) - P_1(J), \text{ for } J = K + \tfrac{1}{2},$$

and

$$\Delta_2 T_2'(J) = R_2(J) - P_2(J), \text{ for } J = K - \tfrac{1}{2},$$

so that,

$$\Delta_2 T'(K) = \tfrac{1}{2}\{\Delta_2 T_1'(J) + \Delta_2 T_2'(J)\}.$$

While in the lower state,

$$\Delta_2 T_1''(J) = R_1(J-1) - P_1(J+1), \text{ for } J = K + \tfrac{1}{2},$$

and

$$\Delta_2 T_2''(J) = R_2(J-1) - P_2(J+1), \text{ for } J = K - \tfrac{1}{2},$$

so that,

$$\Delta_2 T''(K) = \tfrac{1}{2}\{\Delta_2 T_1''(J) + \Delta_2 T_2''(J)\}.$$

The combination differences, $\Delta_2 T(K)$, can be expressed thus:

$$\Delta_2 T(K) = 4B_v(K + \tfrac{1}{2}) + 8D_v(K + \tfrac{1}{2})(K^2 + K + 1),$$

where B_v and D_v depend upon v.

§ 6. ROTATIONAL CONSTANTS

To calculate B_v and D_v corresponding to a given vibrational state the mean values of the combination differences $\Delta_2 T(K)$ have been taken in cases where more than one datum were available for a particular pair of rotational levels. Then by the method of successive approximation the following constants were evaluated:

$$\begin{array}{ll}
B_e' = 0{\cdot}33485 \text{ cm}^{-1} & B_e'' = 0{\cdot}38760 \text{ cm}^{-1} \\
B_0' = 0{\cdot}33350 & B_0'' = 0{\cdot}38640 \\
B_1' = 0{\cdot}33080 & B_1'' = 0{\cdot}38400 \\
\alpha' = 0{\cdot}0027 & \alpha'' = 0{\cdot}0024 \\
D_e' \doteq D_v' = 2{\cdot}01 \times 10^{-7} & D_e'' \doteq D_v'' = 2{\cdot}27 \times 10^{-7} \\
r_e' = 2{\cdot}029 \times 10^{-8} \text{ cm.} & r_e'' = 1{\cdot}886 \times 10^{-8} \text{ cm.} \\
I_e' = 82{\cdot}67 \times 10^{-40} \text{ gm.cm}^2 & I_e'' = 71{\cdot}42 \times 10^{-40} \text{ gm.cm}^2
\end{array}$$

With the values of the molecular constants now available one can check the correctness of the rotational analysis of the bands as follows. On substitution of the values of B_e and D_e in the theoretical relation

$$\omega_e^2 = -4B_e^3/D_e,$$

it is found that $\omega_e' = 864{\cdot}4 \text{ cm}^{-1}$ and $\omega_e'' = 1012{\cdot}9 \text{ cm}^{-1}$. These values are in sufficient close agreement with those obtained from the vibrational analysis of the bands.

Table 2. (1, 0) band, $\lambda\lambda$5469·30 (R), 5470·02 (Q)

K''	R	P	$\Delta_2T'(K)$ $v'=1$	$\Delta_2T''(K)$ $v''=0$
4	18278·81	18270·70	—	—
5	—	69·49	—	—
6	—	68·19	—	—
7	—	66·82	—	—
8	—	65·32	—	—
9	—	63·68	—	—
10	—	61·98	—	—
11	—	60·18	—	—
12	—	58·25	—	—
13	—	56·18	—	—
14	—	53·98	—	—
15	—	51·65	—	—
16	18270·70 70·23	49·05	21·42	—
17	69·49 68·92	46·32	22·88	26·94
18	68·19 67·56	43·52	24·35	28·50
19	66·82 66·05	40·79 40·62	25·73	30·11
20	65·32 64·45	37·90 37·63	27·12	31·73
21	63·68 62·78	34·87 34·54	28·53	33·33
22	61·98 61·02	31·77 31·34	29·95	34·90
23	60·18 59·15	28·59 28·08	31·33	36·51
24	58·15 57·15	25·29 24·70	32·71	38·14
25	56·18 55·01	21·87 21·17	34·07	39·75
26	53·98 52·74	18·35 17·56	35·41	41·25
27	51·65 50·38	14·74 13·94	36·67	42·78
28	49·22 47·92	10·99 10·18	37·99	44·29
29	46·72 45·38	07·14 06·31	39·33	45·84
30	44·12 42·72	03·16 02·30	40·69	47·40
31	41·42 39·90	18199·12 98·19	42·01	48·96
32	38·56 37·00	94·98 93·95	43·32	50·47
33	35·63 34·00	90·77 89·61	44·62	52·03
34	32·54 30·84	86·40 85·11	45·94	53·62
35	18229·35 27·59	18181·87 80·52	47·28	55·14
36	26·06 24·23	77·28 75·82	48·59	56·53
37	22·67 20·78	72·85 71·03	49·78	58·00
38	19·18 17·22	68·13 66·15	51·06	59·50
39	15·57 13·54	63·24 61·20	52·33	61·05
40	11·85 09·73	58·19 56·12	53·64	62·59
41	08·03 05·81	52·95 50·97	54·96	64·27
42	04·09 01·80	47·61 45·44	56·42	65·95
43	00·05 18197·69	42·08 39·87	57·90	67·33
44	95·90 93·49	36·85 34·38	59·08	68·70
45	91·73 89·18	31·58 28·76	60·28	70·37
46	87·56 84·75	25·70 22·94	61·83	71·90
47	82·33 80·22	19·98 17·13	63·22	73·51
48	79·09 75·59	14·17 11·12	64·70	75·10
49	74·67 70·87	08·30 05·02	66·11	76·87
50	70·21 66·05	02·13 18098·82	67·66	78·54
51	—	95·87 92·60	—	80·13
52	—	89·75 86·25	—	—
53	—	83·41 80·07	—	—
54	—	77·06 73·47	—	—
55	—	70·72 67·07	—	—
56	—	64·29 60·57	—	—
57	—	57·93 53·88	—	—
58	—	51·44	—	—

Table 3. (0, 0) band, λλ5735·66 (*R*), 5737·53 (*Q*)

K''	R	P	$\Delta_2 T'(K)$ $v'=1$	$\Delta_2 T''(K)$ $v''=0$
5	17426·92	17417·85	—	—
6	—	16·69	—	—
7	—	15·39	—	—
8	—	13·99	—	—
9	—	12·48	—	—
10	—	10·87	—	—
11	—	09·11	—	—
12	—	05·38	—	—
13	—	03·32	—	—
14	—	01·14	—	—
15	—	17398·87	—	—
16	17418·45	96·48	21·97	—
17	17·33	94·00	23·33	27·08
18	16·06	91·37	24·69	28·69
19	14·69	88·64	26·05	30·14
20	13·21	85·92	27·29	31·81
21	12·48 11·60	83·47 82·99	28·81	33·35
22	10·87 09·90	80·54 79·97	30·13	34·85
23	09·11 08·17	71·49 76·89	31·45	36·37
24	07·32 06·29	74·32 73·71	32·79	37·98
25	05·38 04·32	70·97 70·36	34·18	39·55
26	03·32 02·20	67·62 66·89	35·50	41·14
27	01·14 17399·99	64·09 63·33	36·85	42·68
28	98·87 97·69	60·47 59·69	38·20	44·18
29	96·48 95·30	56·79 55·98	39·51	45·68
30	94·00 92·79	52·97 52·24	40·79	47·14
31	91·37 90·19	49·08 48·42	42·03	48·61
32	88·64 87·46	45·11 44·45	43·27	50·10
33	85·80 84·62	41·05 40·32	44·52	51·59
34	82·83 81·66	36·84 36·08	45·79	53·11
35	79·66 78·58	32·48 31·72	47·02	54·63
36	76·52 75·41	28·00 27·22	48·35	56·12
37	73·23 72·14	23·44 22·57	49·68	57·67
38	69·85 68·79	18·79 17·80	51·02	59·20
39	66·38 65·32	14·05 12·91	52·37	60·76
40	62·82 61·80	09·22 07·90	53·75	62·32
41	59·15 58·12	04·28 02·78	55·10	63·78
42	17355·35 54·32	17299·25 97·81	56·30	65·21
43	51·46 50·44	94·10 92·75	57·52	66·68
44	47·52 46·46	88·84 87·46	58·84	68·20
45	43·48 42·37	83·48 82·02	60·17	69·76
46	39·30 38·19	78·02 76·44	61·51	71·32
47	35·03 33·86	72·41 70·80	62·84	72·94
48	30·73 29·41	66·53 65·07	64·27	74·64
49	26·29 24·88	60·36 59·25	65·79	76·23
50	— 20·23	54·16 53·33	66·90	77·88
51	— 15·52	48·12 47·29	68·23	79·07
52	— 10·69	41·99 41·16	69·53	80·63
53	— 05·74	35·75 34·89	70·85	82·16
54	— 00·71	29·48 28·53	72·18	83·68
55	— 17295·53	23·13 22·06	73·47	85·20
56	— 90·27	16·66 15·51	74·76	86·69
57	— 84·92	10·20 08·84	76·08	88·24
58	— 79·48	03·77 02·03	77·45	89·81
59	— 73·90	17197·38 95·11	78·79	91·41
60	— 68·20	91·32 88·07	80·13	92·98
61	— 62·42	84·48 80·92	82·83	95·95
62	— 56·55	77·73 73·72	84·12	97·48
63	— 50·59	71·45 66·47	85·45	—
64	— 44·52	65·23 59·07	—	—
65	38·27	—	—	—
66	32·01	—	—	—
67	25·62	—	—	—
68	19·13	—	—	—
69	12·56	—	—	—
70	06·26	—	—	—
71	17199·78	—	—	—
72	93·45	—	—	—
73	86·89	—	—	—
74	80·33	—	—	—

Table 4. (0, 1) band, λλ6086·44 (*R*), 6087·49 (*Q*)

K''	*R*	*P*	$\Delta_2T'(K)$ $v'=0$	$\Delta_2T''(K)$ $v''=1$
3	—	16417·90	—	—
4	—	16·79	—	—
5	—	15·56	—	—
6	16425·43	14·21	—	—
7	—	12·78	—	—
8	—	11·30	—	—
9	—	09·76	—	—
10	—	08·12	—	—
11	—	06·37	—	—
12	—	04·54	—	—
13	—	02·63	—	—
14	—	00·56	—	—
15	—	16398·38	—	—
16	16418·52 17·90	96·12	22·09	—
17	17·42 16·79	93·78	23·33	26·90
18	16·18 15·56	91·31	24·56	28·63
19	14·80 14·21	88·48	26·03	30·11
20	13·37 12·78	85·76	27·32	31·56
21	11·89 11·30	82·95 82·47	28·88	33·27
22	10·36 09·76	80·08 79·52	30·26	34·76
23	08·79 08·12	77·15 76·51	31·62	36·29
24	07·12 06·37	74·10 73·43	32·97	37·85
25	05·37 04·54	70·96 70·24	34·35	39·39
26	03·46 02·63	67·72 66·99	35·69	40·94
27	01·47 00·56	64·39 63·64	37·00	42·47
28	16399·37 98·38	60·98 60·16	38·30	43·99
29	97·20 96·12	57·46 56·58	39·64	45·50
30	94·91 93·78	53·84 52·91	40·97	47·00
31	92·52 91·31	50·13 49·19	42·25	48·53
32	90·07 88·78	46·31 45·32	43·61	50·04
33	87·49 86·17	42·38 41·36	44·96	51·59
34	84·78 83·46	38·37 37·30	46·28	53·09
35	82·01 80·64	34·31 33·17	47·58	54·58
36	79·12 77·69	30·15 28·93	48·86	56·09
37	76·14 74·63	25·86 24·61	50·15	57·59
38	73·05 71·47	21·46 20·16	51·45	59·09
39	69·86 68·23	16·96 16·53	52·75	60·56
40	66·62 64·90	12·40 11·00	54·06	62·05
41	63·27 61·50	07·69 06·29	55·39	63·58
42	16359·82 58·08	16302·90 01·45	56·77	65·12
43	56·25 54·54	16298·02 96·50	58·13	66·68
44	52·56 50·85	93·05 91·48	59·44	68·23
45	48·82 47·03	87·96 86·36	60·76	69·74
46	44·97 43·10	82·78 81·14	62·07	71·27
47	40·99 39·09	77·48 75·83	63·38	72·79
48	36·93 35·01	72·05 70·43	64·73	74·31
49	32·79 30·82	66·52 64·93	66·08	75·87
50	28·55 26·05	60·88 59·32	67·42	77·43
51	24·21 22·07	55·14 53·61	68·76	78·69
52	19·76 17·57	49·86 47·80	69·83	80·17
53	15·20 12·97	44·00 41·94	71·11	81·68
54	10·55 08·26	38·04 35·93	72·42	83·16
55	05·81 03·44	32·00 29·84	73·70	84·65
56	00·97 16298·55	25·86 23·65	75·00	86·12
57	96·03 93·58	19·62 17·38	76·30	87·62
58	90·98 88·49	13·28 11·00	77·59	89·10
59	85·86 83·31	06·87 04·53	78·88	—
60	80·61 78·04	—	—	—
61	75·30 72·63	—	—	—
62	69·88 67·20	—	—	—
63	64·37 61·49	—	—	—
64	58·74 55·83	—	—	—
65	53·03 50·23	—	—	—
66	47·25 44·40	—	—	—
67	41·36 38·46	—	—	—
68	35·40 32·45	—	—	—
69	29·37 26·36	—	—	—
70	— 20·18	—	—	—
71	— 13·92	—	—	—
72	— 07·60	—	—	—

A further check is afforded by Birge's rule:

$$2x_e B_e/\alpha = 1{\cdot}4 \pm 0{\cdot}2.$$

For these bands substitution of the proper values shows that

$$2x_e' B_e'/\alpha' = 1{\cdot}55 \quad \text{and} \quad 2x_e'' B_e''/\alpha'' = 1{\cdot}56.$$

It should be noted that the nuclear separation of VO thus obtained is in fair conformity with the value predicted by Mecke(6), who plotted the nuclear separation of all known hydrides of various elements against the respective atomic number of the element. He found that r_e of the hydrides of the elements between He and Ne, Ne and A, A and Kr lie respectively on three widely spaced lines. The nuclear separation of vanadium hydride, according to this graph, should be about $1{\cdot}87 \times 10^{-8}$ cm. The value of r_e for vanadium oxide as estimated by comparing the values of the nuclear separation of the hydrides with those of the oxides is estimated to lie between $1{\cdot}85$ and $2{\cdot}35 \times 10^{-8}$ cm.

§ 7. ELECTRONIC TRANSITION

The analysis reveals that each band consists of two *R* and two *P* branches and is, therefore, due to a transition between two similar doublet states. Even at high rotational quantum numbers none of these branches shows any splitting that could be attributed to Λ-type doubling. Thus the possibility of a $^2\Pi \rightarrow {}^2\Pi$ transition is excluded. Although we cannot ascertain with certainty the number of missing lines in the *P* and *R* series owing to the superposition of the branches and want of resolution in the neighbourhood of the origins of the bands, still the presence of a short and strong *Q* branch rules out $^2\Sigma \rightarrow {}^2\Sigma$(7) leaving $^2\Delta \rightarrow {}^2\Delta$ as the only probable transition concerned.

Further, such a transition is to be expected from the consideration of the molecular electronic states. Assuming that the lower electronic state of VO is derived from a normal V atom and a normal O atom respectively in $3d^3 4s^2 (^4F)$ and $2s^2 2p^4 (^3P)$ states, the possible molecular states are only those with $\Lambda = 1, 3, 4$ and $S = \frac{1}{2}, 1\frac{1}{2}, 2\frac{1}{2}$. Hence the probability of a $^2\Sigma$ or $^2\Pi$ as the ground state of the molecule is excluded. This leads one to suggest that $^2\Delta \rightarrow {}^2\Delta$ is the transition concerned in the emission of the band system, although the criteria of missing lines and intensity relation between the branches for low rotational quantum numbers would have settled this question more definitely.

8. POTENTIAL-ENERGY CURVES AND PRODUCTS OF DISSOCIATION

The actual dissociation energies corresponding to the upper and lower electronic states of the molecule are calculated from the relation

$$\mathbf{D} = \omega_e^2/4x_e\omega_e - \tfrac{1}{2}\omega_e,$$

and it is found that $\mathbf{D}' = 4{\cdot}20$ volts and $\mathbf{D}'' = 6{\cdot}38$ volts. Now by means of the relation given below we can ascertain the electronic states of one of the products of dissociation by calculating E_{atom} from the observed $E_{\text{mol.}}$ and the values of $\mathbf{D}'$ and $\mathbf{D}''$. Assuming as is usually done that the oxygen atom remains in its normal state (^{3}P) in both the excited and unexcited states of the molecule, the electronic state of vanadium atom when dissociated in the excited state can be derived. We know that

$$E_{\text{mol.}}+\mathbf{D}'=\mathbf{D}''+E_{\text{atom}},$$

i.e.

$$E_{\text{atom}}=E_{\text{mol.}}+(\mathbf{D}'-\mathbf{D}'').$$

In the band system under consideration, $E_{\text{mol.}}$ corresponding to the origin of the (0, 0) band is 2·15 V., so that $E_{\text{atom}}=-0{\cdot}03$ V. Considering the limits of extrapolation error in the values of $\mathbf{D}'$ and $\mathbf{D}''$, we may take E_{atom} as 0. It is therefore clear that both the $^2\Delta$ states of VO dissociate into the same products, namely, the unexcited atoms of vanadium and oxygen, V (^{4}F) + O (^{3}P).

Similar conclusions have been drawn by Mulliken$^{(8)}$ in the case of N_2^+ negative system involving a $^2\Sigma\rightarrow{}^2\Sigma$ transition and few other isoelectronic diatomic molecules. For N_2^+ Mulliken also predicts the existence of a band system, probably lying in the far red, from an analogy with these molecules. In the case of VO one is led to suspect that the bands recorded by Ferguson in the far-red region of the spectrum might be a part of another system for the molecule.

It is interesting to note that for the oxides of the elements of the first transition group, E_{atom} calculated from the data of the blue bands (system I) of ScO, of the blue-green bands of TiO, and of the band systems known for CrO and MnO, is found to be negligibly small in each case. These values are shown in table 5.

Table 5

	$E_{\text{mol.}}$	$\mathbf{D}'$	$\mathbf{D}''$	E_{atom}
ScO	2·5	4·6	7·5	−0·4
TiO	2·4	4·8	6·9	+0·3
VO	2·2	4·2	6·4	0·0
CrO	2·0	1·9	3·8	+0·1
MnO	2·2	1·7	4·4	−0·5

With the data now available for ω_e, r_e and $\mathbf{D}_e$ it is possible to construct, after Morse, the $U(r)$ curves for the excited and unexcited states of the vanadium oxide molecule. These curves are given in figure 2. They meet one another for values of r much greater than r_e and represent the case of a pair of molecular states which dissociate into the same atomic states.

§ 9. ACKNOWLEDGMENTS

The author wishes to express sincere thanks to Prof. P. N. Ghosh for his encouragement during the course of the investigation and to Prof. R. W. Wood for kindly supplying the 21-ft. grating without which the present work could not have been carried out.

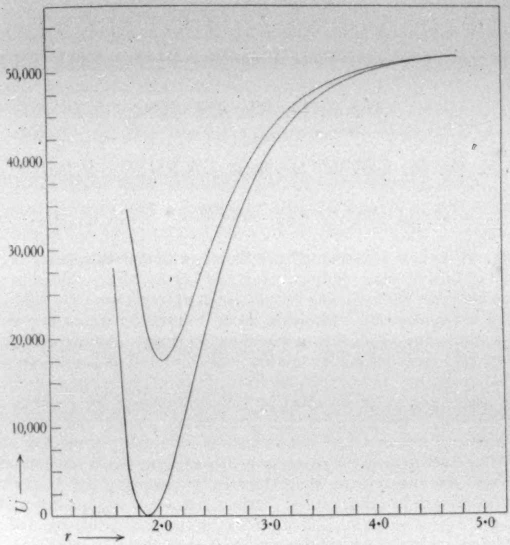


Figure 2. Curves relating *U* with *r*.

REFERENCES

(1) Demarçay, E. *Spectres électriques*, Paris (1895).
(2) Kayser, H. *Handbuch der Spectroscopie*, **6**, 786 (1912).
(3) Ferguson, W. F. C. *Bur. Stand. J. Res.*, Wash., **8**, 382 (1932).
(4) Mecke, R. and Guillery, M. *Phys. Z.* **28**, 514 (1927).
(5) Christy, A. *Phys. Rev.* **33**, 701 (1929).
(6) Mecke, R. *Bandenspektra und ihre Bedeutung für die Chemie*, p. 26 (1929).
(7) Ghosh, C. S. *Nature*, Lond., **132**, 318 (1933).
(8) Mulliken, R. S. *Rev. mod. Phys.* **4**, 48 (1932).

771.36

NEW METHOD OF MEASURING THE TIME AND EFFICIENCY OF PHOTOGRAPHIC SHUTTERS*

By E. D. EYLES, B.Sc.

AND

E. W. H. SELWYN, B.Sc., A.R.C.S., F.Inst.P.

Received January 17, 1935. *Read in title March* 15, 1935.

ABSTRACT. Apparatus is described for obtaining the curves relating area with time in photographic between-lens shutters. A set of separated and illuminated apertures fitted with neutral glass filters passing certain standard fractions of the light is imaged by a lens on film wrapped round a revolving drum, the shutter acting as a diaphragm for the lens. The film is thus exposed in bands along its length. An exposure is made, with the shutter set fully open, for one revolution of the drum. The above apertures are then replaced by a second set without neutral filters, each of these apertures fitting exactly between the spaces previously occupied by two apertures of the previous set. An instantaneous exposure of the shutter is then made during another revolution of the drum. During the whole of this exposure a time scale is impressed on the film by subsidiary apparatus. After development the points at which adjacent bands have the same density are found. These give the times at which the area of opening of the shutter was equal to the standard fractions of its full aperture. Various features of the apparatus are discussed, and particulars are given of the method of calibrating the neutral filters and of tests carried out on shutters for which curves relating area with time could be calculated.

---

§ 1. INTRODUCTION

The ideal photographic shutter opens to its full aperture infinitely quickly, stays fully open for a predetermined period, and then closes infinitely
A, T quickly. If A is the area of the full aperture and T the time of opening, the total amount of light admitted by the shutter is proportional to AT. In actual fact a shutter takes some time to open and some time to close, and the total amount
a, t of light admitted is proportional to $\int_0^T a\,dt$, where a is the area open at a time t, measured from the commencement of opening, and T the total time from beginning to end of the exposure. If the shutter has a full aperture A, the quantity $\frac{1}{AT}\int_0^T a\,dt$ is termed the efficiency.

The first measurements of efficiency appear to have been made by Abney in 1892⁽¹⁾. A more accurate method following the same principles was published in 1909⁽²⁾ from the National Physical Laboratory. This gives a record of the length of a line, across the aperture of the shutter, which is visible at any instant during the exposure. By independent experiments is obtained the relation between the area of opening of the shutter and the length of line visible, which enables the relation between area of opening and time to be plotted for the shutter. One of these operations was eliminated in apparatus devised in the Kodak Research Laboratories⁽³⁾ which gave a high-speed kinematographic record of the position of the shutter blades at short, equally-spaced intervals of time. For accurate estimation of the efficiency these methods require the measurement, with a planimeter or otherwise, of the area uncovered by the shutter blades at recorded instants.

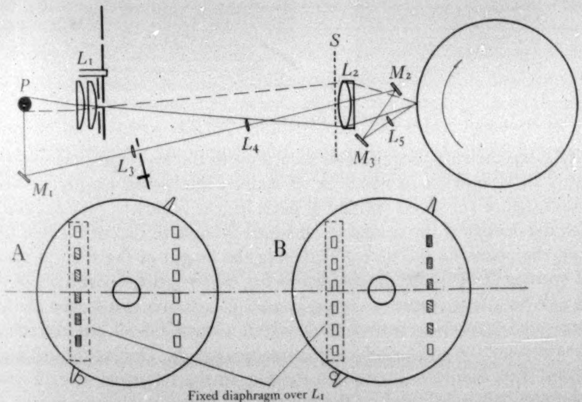
---

* Communication H 544 from the Kodak Research Laboratories.

The apparatus now to be described gives a photographic record of the time at which the area of the shutter uncovered by the blades amounts to certain fractions of the aperture when fully open; that is, it provides directly a certain number of points on the area-and-time curve of the shutter. It was built specifically for between-lens shutters with a full aperture of diameter not greater than 25 mm. and exposure times of from 1·0 to 0·004 sec.

§ 2. DESCRIPTION OF APPARATUS

An image of the incandescent ball P of a 500-candle-power pointolite lamp is formed by the condenser lens L_1 at the plane S of the shutter blades, figure 1. Just in front of the condenser lens is a diaphragm which may be rotated into either of two positions. In one of these positions it forms a column of equal apertures at equal intervals, coming within the fixed rectangular aperture over the condenser lens indicated by the broken lines, figure 1 A. In the other position a second column of similar apertures fits exactly between the spaces previously occupied by the first set of apertures, figure 1 B. All but one of the apertures of the first set have



Fixed diaphragm over L_1

Figure 1. Diagram of apparatus.

neutral glass filters fitted so that the transmissions of the apertures are 1·0, 0·9, 0·7, 0·5, 0·3 and 0·1 approximately. An image of the apertures is formed by the lens L_2, placed close to the plane S, on the surface of a strip of kinematograph film wrapped round a drum which can be steadily rotated.

An exposure is made to the first set of apertures for one revolution of the drum with the shutter under test open at full aperture. The film is thus, owing to its motion, exposed in bands corresponding to the image of the apertures at intensities 1·0, 0·9, 0·7, 0·5, 0·3 and 0·1 of the intensity transmitted by the shutter at full aperture; see figure 2*a*. Then the diaphragm in front of the condenser is rotated into the other position, and an exposure is made with the shutter set at the instantaneous exposure being measured. This results in another exposure in bands, corresponding to the image of the second set of apertures, figure 2*b*, but in this case the exposures in all the bands are the same but increase and then diminish in intensity along the length of the film according to the area-and-time curve of the shutter. This second set of bands is, of course, interposed exactly between the first bands, figure 2*c*. When the film is developed it is easy to pick out the points at which the density of the second set of bands is equal to that of the first set: that is to say, the points on the time scale (along the length of the film) at which the area of opening of the shutter amounted to 1·0, 0·9, 0·7, 0·5, 0·3 and 0·1 of the full aperture. When these points are being located it is convenient to rest the film on an illuminated opal glass, with a mask which screens off all but the immediate neighbourhood of the point at which the densities match. During the instantaneous exposure a time scale is impressed on the film in the following way. Light from the pointolite ball is reflected by the mirror M_1, figure 1, along a line at a small angle to the axis of the apparatus and crossing the axis at the plane S. The light is made to strike the film at the proper place on the drum by the mirrors M_2 and M_3. This double reflection is necessary to avoid interference by the lenses and mirrors with the light travelling along the axis of the apparatus. The lens L_3 forms an image of the pointolite ball on L_4, and L_4 forms a reduced image of the aperture over L_3 at the central point in the plane S, and this small image is again reproduced

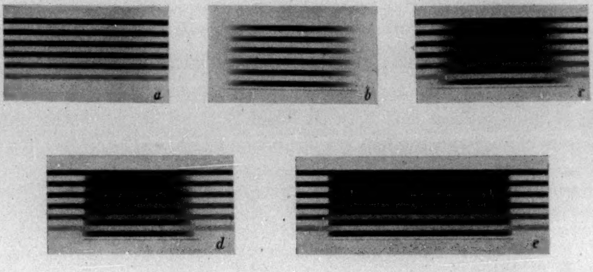


Figure 2. Examples of records.

on a still smaller scale on the film by the lens L_3. In this way sufficient light to give a record on the film is transmitted with exceedingly small openings of the shutter. The principle of sending a very small but intense beam of light through the centre of the shutter aperture is that used in the National Physical Laboratory apparatus(2). Immediately in front of the lens L_3 is mounted a disc with evenly-spaced holes cut near the periphery, and made to revolve at high speed by a small motor synchronized with controlled-frequency alternating current by apparatus of the phonic-wheel type, such as is commonly used in television receivers. The light is interrupted in this way 1995 times a second, which is as near as could be obtained to the desired 2000 times a second, with the a.-c. frequency and apparatus available. For the total duration of opening of the shutter, therefore, flashes of light are being sent through the shutter, and are recorded on the developed film as short, slightly inclined black lines, spaced at intervals along the film corresponding nearly enough to 0·0005 second.

§ 3. UNIFORMITY OF ILLUMINATION OF SHUTTER

It is obvious that for accurate results the intensity of illumination should be uniform over the area of the full aperture of the shutter. Experiment has shown that the pointolite ball is, in itself, satisfactory in this respect. But the condenser lens is not corrected and this causes some difficulty. Chromatic aberration and coma do not cause any serious trouble since the brightness of the image is the same as if there were no such aberration, except for a rim round the edge of the image of the same order of dimensions as the size of the diffuse patch which is formed, in consequence of these aberrations, as the image of a point source. But the spherical aberration is of more consequence: as a result of it, the images given by the different apertures do not overlap exactly. Sufficient overlap exists, however, to give an area 25 mm. in diameter illuminated uniformly by the light from any small element of the condenser. The fact that the brightness of the image produced by each small element of the condenser is not the same as that produced by another element is immaterial, as will be shown later.

§ 4. METHOD OF INTENSITY-COMPARISON

The valuable features of the method using juxtaposed bands of varying density in the photographic comparison of intensities have been pointed out in considerable detail by L. A. Jones(4). If these features are to be retained properly it is very necessary that the strips be in as exact contact as possible. Therefore the diaphragm and apertures were made very carefully, so that one set of apertures fits precisely in between the other set. Since the areas on the film at which the comparison is made are small and very close together it is not necessary to make special efforts to secure uniform development. For the same reason accurate results are not dependent on the use of sensitive material with more than ordinary uniformity. Again, since the comparison of densities is, in effect, made along the line of contact of two bands, variation in the illumination from different points on the condenser lens is without effect provided that there is no sudden variation close to the dividing line between two bands or apertures.

The average illumination on the film depends upon the size of the full aperture of the shutter being measured, and in order to keep the density of the record at a convenient level a variable resistance is inserted into the supply feeding the pointolite lamp, so that for large-aperture shutters the intensity of the pointolite ball may be reduced. In this way the same average density may be obtained with a shutter having an aperture 25 mm. in diameter as with one having an aperture 7 mm. in diameter.

At first sight an attempt to obtain bands differing in exposure by reducing the width of the apertures to 0·9, 0·7, 0·5, 0·3 and 0·1 of the maximum is attractive. But it is difficult to do so when, as will be seen later, the maximum width is for good reason only 2 mm. Moreover, apart from this difficulty, the method is theoretically unsound owing to the influence of reciprocity failure.* Reduction of the widths to the above values was in fact found not to give the required reduction in exposure. Even if the aperture-widths were reduced empirically to simulate reduction in intensity to 0·9, 0·7, 0·5, 0·3 and 0·1 of the maximum, there still remained the possibility of the introduction of errors unless the intensity of illumination on the film and the speed of the drum were always the same. The use of neutral filters permits the exposures to be kept of constant duration and thus avoids any difficulties due to reciprocity failure.

Cine positive film is used for the record because it is easy to handle, an orange light being safe enough with it, while its speed is adequate. The contrast is high so that differences in intensity of illumination are reproduced strongly, and determination of the position of equal densities is made easier. Since this material is not colour-sensitized and since much of the ultraviolet radiation from the lamp is lost by absorption in the lenses of the apparatus, the effective sensitivity is restricted to a relatively narrow region of the spectrum, and perfect neutrality of the filters is therefore unnecessary. For the same reason a change in the colour temperature of the pointolite lamp is practically without effect on the effective transmission of the filters. In order to keep the intensity as steady as possible the lamp is run off accumulators: the variation in intensity from one exposure to the next is negligible.

The gramophone motor which drives the drum is driven by controlled alternating current and is found to keep within ± 1 per cent of its average speed. Stroboscopic examination shows that the drum is not subject to sudden fluctuations in speed, such as would average out over a long period and not be detected when the exposure is timed with a stop watch. The position of the drum at any instant during its rotation is shown by an index on the lid of the box enclosing the drum. This index is loosely coupled to the drum (so that the lid may be lifted up for wrapping film round the drum) and rotates in synchronism with it. With the aid of the index it is possible to start and to end the exposure to the apertures fitted with filters at the point on the drum where the film is joined, and also to time the instantaneous exposure of the shutter so that it falls midway between the ends of the film. After practice it is, in fact, quite possible to space out as many as four instantaneous exposures on one film.

---

* "Reciprocity failure" is the term used to describe the fact that the density produced under given conditions of development is not a function of the product of intensity of illumination and time of exposure alone, or that, in other words, an x times decrease of intensity is not necessarily compensated for by an x times increase in exposure-time.

§ 5. EFFECT OF FINITE WIDTH OF APERTURES

Suppose that a point x on the film, which is moving with a velocity v, passes under the second edge of the image of the aperture at a time t. If the width of this image is Δ, then the exposure commences Δ/v before t and finishes at t. If the intensity of the illumination passed by the shutter at the time t be represented by $f(t)$, the total exposure at x is $\int_{t-\Delta/v}^{t} f(t)\,dt$. On a band where the illumination is constant and equal to I, the total exposure anywhere along the film is $I.\Delta/v$. The points where the densities of the two bands are equal is given by the value of t at which

$$I.\Delta/v=\int_{t-\Delta/v}^{t} f(t)\,dt$$

or

$$I.\Delta/v=f(t-\tau).\Delta/v,$$

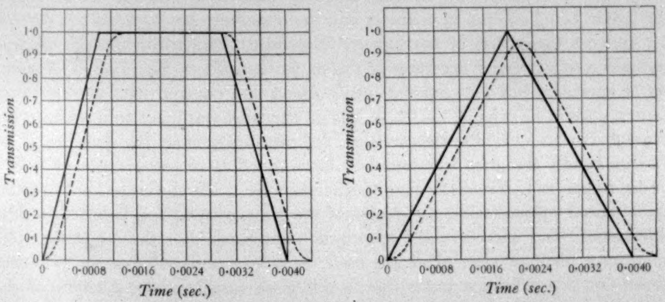


Figure 3. Effect of finite aperture width.

where τ lies between 0 and Δ/v and is such as to make $f(t-\tau)$ equal to the mean value of $f(t)$ over the range $(t-\Delta/v)$ to t. The effect of using apertures of finite width is therefore to displace the curve bodily along the time axis by an amount roughly equal to $\Delta/2v$ where the curve has no sharp curvature over the interval Δ/v, and to round off all sharp bends. The small dot of light which traces out the time scale is therefore made to fall on the vertical centre line of the image of the column of apertures. The amount of distortion of the curve caused by rounding off sharp bends increases relatively at the faster shutter-speeds. However, in the present case Δ is 0·7 mm. and $1/v$ is equal to 0·0006 sec./mm. on the average, so that even at a total time of opening of 0·004 sec. the distortion is not very serious. This is shown by figure 3, where hypothetical curves, as they would be recorded with an indefinitely small value of Δ (or infinite value of v), are shown by the full lines, and the curves as they would be recorded under the conditions stated above are shown by the broken lines.

§ 6. CALIBRATION AND TESTS

The uniformity of illumination over the shutter was tested by exposing process film* placed at the plane S, only one of the apertures being used over the condenser lens. The density-distribution over the resulting image was measured after careful development and fixation, and it was found that over an area 25 mm. in diameter there was no variation in intensity greater than ± 5 per cent, whichever aperture was used. A variation of this amount is practically negligible for the present purpose. Indirect proof of the uniformity of illumination is also provided by the tests on shutters for which the curves relating time with area are known.

As the filters had not the same transmission at all wave-lengths, it was necessary to reduce them to the correct thickness empirically. This was done with the apparatus in the following way, and it may be noted that if any variation of intensity were centrally symmetrical, as variations are of course most likely to be, the method of calibration would tend to reduce its effect on the final result when shutters are being tested. A set of circular apertures was made to fit into the shutter-holder, the largest being 20 mm. in diameter and the remainder having areas 0·9, 0·7, 0·5, 0·3 and 0·1 times that of the largest. Exposures were made through them in succession to successive apertures of the set without filters, figure 1 B. The second set of apertures was then rotated into position and covered with a piece of the neutral glass in its original thickness, and an exposure was made to these apertures through the 20-mm. circular aperture. By measurement of the densities of the various bands, the glass was found to have an effective transmission of a little under 10 per cent. From this figure estimates were made of the thicknesses giving the required transmissions, and pieces of the glass, ground and polished to these thicknesses, were placed in the appropriate apertures. For the 0·9-transmission aperture a piece of plate glass was used, as it was expected that the combined absorption and reflection losses due to this glass would reduce the transmission very nearly to the value in question. The previously described operations of exposure through the apertures in the shutter-holder were again carried out. When the densities of the bands on the developed film had been measured, the densities of the bands exposed through the set of shutter apertures were plotted against the relative areas of the shutter apertures. The transmissions of the filters were then read off this curve from the value of the densities of the bands exposed through the filters and the 20-mm. shutter aperture. This is allowable if the filters are in their final order on the diaphragm and if the exposures through the shutter apertures are made so that bands of the same nominal intensity (one through the filter and one through the corresponding shutter aperture) are adjacent. For then by measuring the densities over small areas (1 mm. in diameter) close to the line dividing the strips, the effect of different light transmission through different parts of the condenser is almost eliminated. A nearer approximation to the required thickness is thus obtained. It has also been found in this way that alteration of the intensity of the pointolite lamp (by lowering of its temperature) is without measurable influence on the effective transmission of the filters.

---

* Process film was used since it will, by its high contrast, exaggerate any non-uniformity of illumination.

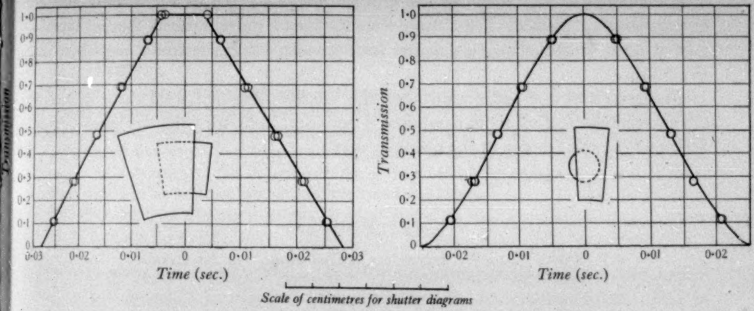


Figure 4. Experiments on shutters with known efficiency curves.

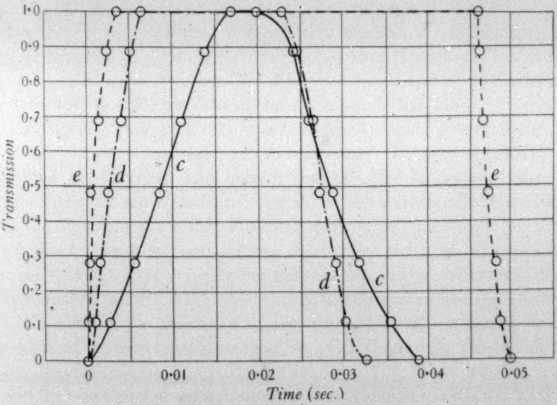


Figure 5. Three commercial shutters.

The transmissions of the filters as finally used were measured in the above way and found to be

$$0{\cdot}11,\quad 0{\cdot}28,\quad 0{\cdot}48_5,\quad 0{\cdot}69,\quad 0{\cdot}89.$$

It is regarded as very unlikely that any of these values is in error by more than ±0·01.

Tests have been carried out on two shutters for which the curves relating area with time could be calculated. These consisted of sector apertures rotated at constant speed by a synchronous motor over either a sector-shaped or a circular fixed aperture. The shapes and sizes of the shutters are indicated in figure 4, together with the calculated curves relating area with time and the observed points. Two experiments were made on the first shutter and three on the second, and all observations are plotted. From these tests it appears to be unlikely that an experimental error greater than 0·02 in the transmission would be found anywhere on the curve.

In figure 2 at *c*, *d* and *e* are shown the records obtained from three different commercial shutters all set to a nominal $\frac{1}{25}$ sec., and in figure 5 the efficiency curves derived from the records in the ordinary way.

REFERENCES

(1) Abney, W. de W. *Instruction in Photography*, p. 106 (1900).
(2) Campbell, A. and Smith, T. *Proc. phys. Soc.* **22**, 788 (1907–9).
(3) Nutting, P. G. *Photogr. J.* **63**, 394 (1916).
(4) Jones, L. A. *J. opt. Soc. Amer.* **10**, 561 (1925).

55.034.1

TIME AND AMPLITUDE RELATIONS IN SEISMOLOGY

By HAROLD JEFFREYS, F.R.S.

Received November 27, 1934. *Read March* 1, 1935.

ABSTRACT. The paper deals with certain anomalies in the amplitudes of the waves observed in near earthquakes, and their relation to the method of transmission across interfaces. It is suggested that amplitude-observations in seismic prospecting may help to solve these theoretical problems. A rediscussion of the time relations shows that these are so closely related to the structure immediately below the observing station that the form of the interfaces produces no other appreciable effect.

SOME years ago I published a paper⁽¹⁾ on compressional waves in two superposed layers, which appears to have attracted an amount of attention in seismic prospecting that I should never have anticipated. The work described in it was done with the object of finding out whether the possibility of diffraction along an interface would explain the observed variation with distance of amplitude of vibration in near earthquakes. It failed in its purpose in two respects, and I am not yet able to ascertain from the literature whether it succeeds any better in its application to seismic prospecting. Without wishing to appear ungracious, I think that it is desirable to call attention to the outstanding difficulties.

The elementary theory of elastic waves from a local source shows that the amplitude should vary inversely as the distance.* In near earthquakes—those well observed at distances less than a few hundred kilometres—the observed amplitudes diminish with distance, but definitely less rapidly than the inverse square of the distance; the actual variation is irregular, but roughly as the inverse distance.† This would be satisfactory if it were not for two complications. The motion of the ground depends not only on the incident waves, but also on the reflected ones, and for a *P* wave at grazing incidence the effects of these two just cancel.‡ To explain the existence of any ground-motion it seems necessary to take the depth of the focus into consideration; but when we do this and allow for the conditions of reflection we find that a factor depending on the angle of emergence enters, and the result is that the ground movement should vary as the inverse square of the distance.§ This applies to direct waves. For waves that have travelled in a lower layer the angle of emergence is always finite, and this complication does not arise. But when the path in the lower layer is nearly horizontal the loss of energy by reflection in crossing the interfaces is very great. If we adopt the usual theory and suppose the energy to travel along the rays, correcting for this loss, we find that the variation of amplitude with distance disappears at short distances; the greater amplitude in the original wave is cancelled by the greater loss on reflection when the path in the lower layer is nearly flat.* This conclusion again disagrees with the facts, the observed amplitudes being again roughly as the inverse distance.

---

* See reference (2), p. 282.

† See reference (3), p. 400.

‡ See reference (4), p. 327.

§ See reference (4), p. 336.

My paper was intended to find out whether these results would be altered if we allow for diffraction at the interface and ignore the curvatures of the rays and the interface, the difference of which is what matters in the elementary theory. For its original purpose the work was a failure. The surface motion due to the direct wave was still as the inverse square of the distance. There was an indirect wave, which would not exist at all in these conditions on the ordinary theory, but it still disagreed with observation. If the direct wave begins with a finite velocity, the indirect one should begin with a finite acceleration; the shorter periods present in the direct wave should therefore be smoothed out in the indirect one. The acceleration also should vary as the inverse square of the distance. Actually the indirect wave, except for being smaller, is very like the direct one; and short periods are equally prominent.

It seems to me that the amplitude relations are capable of only one interpretation, namely that the angle of emergence of the direct waves is always finite and much the same at all distances, and that the same applies to the angle of emergence of the indirect wave at the interface. In this way the troublesome extra factors are avoided. But this cannot be achieved in accordance with the ordinary law of refraction. We seem to need scattering, such as occurs in the case of light at the surface of ground glass. In near earthquakes this may be due to minor irregularities in structure, and possibly in the form of the interface. How far this applies to the conditions of seismic prospecting I am not in a position to say, but I think that this field of study may give valuable information, since the form of the interface is there ascertainable. For comparison with the results of the study of near earthquakes I should like to know how the amplitudes do in fact vary with distance, whether there is any systematic difference between the direct and indirect waves in respect of the prominence of short periods and sharp onsets, and how these relations vary according as the interface is plane, anticlinal or, if possible, synclinal. In these respects seismic prospecting may make a valuable contribution to pure seismology.

The various theories all give the same results for the times of arrival of the pulses. A little more discussion of the effect that arises when the interface is not horizontal, however, seems to be desirable. In works dealing with seismic prospecting a formally exact formula, which is correct for a plane sloping interface, is usually given. But the time is obviously unaffected by the slope of the interface at any point other than those of entry into the lower medium and emergence from it, and if the interface is anticlinal the mean slope will in general differ from the actual slope at both these points. The exactness of the formula is therefore only apparent. There may be advantages in replacing it by an adaptation of the corresponding formula used in the study of near earthquakes. The time T of transit is $\int ds/c$, taken along the path, where ds is the element of length and c the local velocity. If v is the velocity at the lowest point of the path, x the horizontal distance reached, and e the angle

T
ds, c, v
x, e

---

* See reference (4), p. 340.

of emergence at any point (that is, the angle between the ray and the horizon), we can write this

$$T=\frac{x}{v}+\int\left(\frac{\sec e}{c}-\frac{1}{v}\right)dx \qquad \text{......(1).}$$

But if z is the depth, $dz=\tan e\,dx$, and

$$T=\frac{x}{v}+\int\left(\frac{\sec e}{c}-\frac{1}{v}\right)\cot e\,dz \qquad \text{......(2).}$$

Where the stratification is horizontal we have, by the law of refraction,

$$\cos e=c/v \qquad \text{......(3),}$$

$$T=\frac{x}{v}+\frac{1}{v}\int\left(\frac{v^2}{c^2}-1\right)\left(\frac{v^2}{c^2}-1\right)^{-\frac{1}{2}}dz$$

$$=\frac{x}{v}+\int\left(\frac{1}{c^2}-\frac{1}{v^2}\right)^{\frac{1}{2}}dz \qquad \text{......(4),}$$

the root being taken positive on the downward path and negative on the upward path(5). The coefficient of dz in the integral has been called the "delay depth coefficient" by A. W. Lee(6). For a single upper layer of depth H, with a surface focus, we have

$$T=\frac{x}{v}+2H\left(\frac{1}{c^2}-\frac{1}{v^2}\right)^{\frac{1}{2}} \qquad \text{......(5).}$$

For a sloping interface equation (2) is still exact but (3) is altered, e being changed by a small quantity of the order of the slope of the interface. But it is easy to see that the effect on (4) is negligible. We are still at liberty to consider a ray whose inclination in the upper layers is determined at every point by (3). In the lower layer we must consider it as straight but inclined to the horizon at a small angle α. For this part of the path

$$\frac{\sec e}{c}-\frac{1}{v}=\frac{1}{v}(\sec\alpha-1) \qquad \text{......(6),}$$

and is small, of order α^2/v. Hence to the first order in α the time of the ray is still given by (4), where the integral is taken through the upper layers with their actual thicknesses at the places of entry and emergence.

Such a ray does not strictly satisfy the laws of refraction, because e is altered by quantities of the order of the slope of the interface at the points of entry and emergence. But the time of transit is stationary for small variations of the path, and therefore (4) differs from the true minimum time by small quantities of the order of the squares of the slopes. Thus, finally, if we write

$$T=\frac{x}{v}+\int\left(\frac{1}{c^2}-\frac{1}{v^2}\right)^{\frac{1}{2}}|\,dz\,| \qquad \text{......(7),}$$

where the integral is taken through the upper layers and the root is always taken positive, the error committed is of the order of the squares of the slopes, and in practical conditions is negligible. When the result is stated in this form we see that the delay in transmission due to the overburden depends on the velocities and on the depth of the overburden near the ends of the path, and only to an unimportant extent on anything else.*

a

$, H_1, H_2, H_3$

η

As an example, let us consider explosions at two points O and A at a distance a, and suppose that there is a third point of observation P on the line joining them and distant x from O. The depths at O, A and P are H_1, H_2, H_3. Denote the delay-depth coefficient by η. Then

$$T_{OA}=T_{AO}=\frac{a}{v}+\eta\,(H_1+H_2),$$

$$T_{OP}=\frac{x}{v}+\eta\,(H_1+H_3),$$

$$T_{AP}=\frac{a-x}{v}+\eta\,(H_2+H_3).$$

Hence
$$T_{OP}+T_{AP}-T_{OA}=2\eta H_3,$$

y, H_4

so that H_3, which is what we wish to know, is directly determined. This of course assumes that the velocities are known already. But if we have a further observation point Q distant y from O, where the depth is H_4, we obtain

$$T_{OP}-T_{OQ}=\frac{x-y}{v}+\eta\,(H_3-H_4),$$

$$T_{AQ}-T_{AP}=\frac{x-y}{v}-\eta\,(H_3-H_4),$$

whence v is determined by addition. The velocity in the overburden needs to be determined independently.

This argument naturally adds nothing to the present technique, but it shows that the usual methods are more directly related to the depth to be found than might be supposed from the form in which the subject is usually presented.

---

* I gather from a referee's remarks that a similar point has recently been made by Dr J. H. Jones, but I have not yet seen his paper (7).

REFERENCES

(1) Jeffreys. *Proc. Camb. phil. Soc.* **23**, 472–481 (1926).
(2) Love. *Mathematical Theory of Elasticity* (1906).
(3) Jeffreys. *Mon. Not. R. astr. Soc.* Geoph. Suppl. **1**, 385–402 (1926).
(4) Jeffreys. *Mon. Not. R. astr. Soc.* Geoph. Suppl. **1**, 321–348 (1926).
(5) Wrinch and Jeffreys. *Mon. Not. R. astr. Soc.* Geoph. Suppl. **1**, 20 (1923).
(6) Lee. *Mon. Not. R. astr. Soc.* Geoph. Suppl. **3**, 14 (1932).
(7) J. H. Jones. *Proc. World Petroleum Congress* (1934).

DISCUSSION

Prof. A. O. Rankine. I am glad that the author has given the Society this re-discussion of the basis of the so-called refraction method in seismic prospecting. It is true that recently this method has largely given place to the reflection method, but it retains its importance as giving more information about the nature of underground strata. The author's present contribution will tend to prevent waste of time in seeking formally exact solutions in structural problems which do not really justify the attempt. In particular it provides a means of mapping underground contours of small relief over large areas. As the author mentions, Dr J. H. Jones has at least partially recognized the closeness of the relation between local depth and time of propagation of the indirect disturbance, and has used it to map limestone contours in Persia.

With regard to the author's query as to whether the seismograms obtained in seismic prospecting may serve to elucidate the mechanism of the production of the indirect disturbance, I should say probably not, with the possible exception of those obtained with J. H. Jones's seismometer. The majority of portable seismometers, having been designed primarily to detect the time of the first onset of the disturbance whether direct or indirect, are by no means free from resonance, and the form of the record is not a true representation of the earth's movement. I suggest that the author should, if possible, examine some of the many seismograms Dr Jones has obtained from the point of view mentioned above. I hope that he will eventually be able to give us the true explanation of the phenomena, which are real enough.

541.183.36

ADSORPTION OF GASES ON MERCURY

By R. S. BURDON, University of Adelaide

Communicated by Prof. Kerr Grant, November 12, 1934. *Read in title December* 21, 1934.

ABSTRACT. A mercury surface was formed in the presence of a gas and the gas was pumped off after an interval. The mercury surface was then caused to collapse. This caused an evolution of gas which was measured with a McLeod gauge. For air, hydrogen and carbon dioxide the limiting amount evolved corresponded to a monomolecular layer. This layer appeared to be retained by the surface for a long period. If adsorption beyond a monomolecular layer occurred the subsequent layers were desorbed on evacuation.

§ 1. INTRODUCTION

Some years ago Prof. Kerr Grant suggested to the author the problem of investigating the spreading of liquids, particularly water and aqueous solutions, on the surface of mercury. The discussion of the phenomena observed demands a knowledge of the surface tension of mercury both in air and against its own vapour, as well as of the interfacial tension between mercury and the liquids used. Reference has been made elsewhere to the extreme confusion in the literature regarding both the value of the surface tension of mercury and the action of gases on its surface, and several papers from this laboratory have described work carried out on these subjects(1–8).

Briefly the position was that workers using the method of the sessile drop obtained values of about 440 dyne/cm. for the surface tension of mercury against its own vapour, but for surfaces formed in gases the initial values were about 500 dyne/cm., falling rapidly at first and then more slowly to the value for vacuum or lower. These results were accepted by authors of standard text books. Now the drop-weight method is almost the only other one than can be applied to determine surface tension both in gases and in a vacuum, and experiments made here(2) as well as the very careful work of Harkins(9) showed that this method yielded very nearly the same value for the surface tension in air as in a vacuum. There was thus the problem of the actual value of the surface tension of mercury and also of the behaviour of its surface with respect to gases. By means of an apparatus constructed of fused silica and the method of the sessile drop the surface tension was determined in a high vacuum and also in the presence of air, both dry and moist(8). This yielded essentially the same value for the surface tension in gas and in a vacuum (as was indicated by drop-weight measurements), and showed that differences between these values observed by various workers could not be attributed merely to the presence or absence of the gas.

The question whether gases were really adsorbed by the mercury was attacked by Oliphant$^{(5)}$ who found that drops of mercury formed in an atmosphere of either argon or hydrogen containing as much as 2 per cent of carbon dioxide became almost completely covered with a layer of molecules of carbon dioxide by the time the drops had fallen 1 ft. through the mixture (i.e. within ¼ sec. after formation). Bosworth$^{(7)}$ confirmed this result and extended the measurements to sulphur dioxide and to water vapour. These apparently represent the first direct measurements of the amounts of gas adsorbed at a mercury surface, though Iredale$^{(10)}$ had estimated the adsorption of vapours by applying the Gibbs adsorption equation to measurements of surface tension, and Sissingh and Haak$^{(11)}$ and more recently Herschkowitch$^{(12)}$ have estimated the condensation of gases on the surface from variations in the optical properties of the surface when exposed to the gas. The theory of this last method has been discussed by Tronstad and Feachem$^{(18, 19)}$.

Oliphant's results for carbon dioxide in hydrogen leave three possibilities as to the behaviour of these gases and hydrogen: (i) that carbon dioxide but not hydrogen is adsorbed by the expanding mercury surface; (ii) that hydrogen is adsorbed but that its life on the surface is extremely short compared to that of carbon dioxide; (iii) that hydrogen by itself may be adsorbed and persist on the surface but that the presence of the carbon dioxide actually causes the hydrogen to be desorbed.

Either the first or the second of these assumptions gives a simple picture of the process occurring at the surface but neither of them can be reconciled with the results of the present work or with sundry observations recorded in the literature$^{(13, 14)}$ regarding the reaction of hydrogen and mercury. Bosworth's results indicate that when carbon dioxide and water vapour are each present, the relative amounts present in the adsorbed layer and in the gas are not the same, as would be the case if each kind of molecule were adsorbed on impact and had a considerable life on the mercury surface independently of the other active substance*.

The aim of the present work was to determine whether gas adsorbed on the surface of mercury leaves the surface as the gas pressure above it is reduced to low values, and if it does not, to measure directly the amounts adsorbed by both static and expanding surfaces. Calculation shows that a monomolecular layer of gas over a surface of moderate area involves a quantity of gas sufficient to give a measurable pressure in an apparatus having a volume of a litre or more. Owing to the vapour pressure of the mercury itself it was impossible to use any of the special types of manometer such as ionization or Pirani gauges, and a McLeod gauge giving a reading of 2·3 cm. for a pressure of 0·001 mm. was used.

The original idea was to form a surface some 400 cm^2 in area in a flat chamber in the presence of the gas, to exhaust to a low pressure, and then to measure the gas evolved when the surface was destroyed by allowing the mercury to flow out through a narrow tube at the bottom. It was necessary to be able to form the surface and to run the mercury out with the apparatus exhausted without the use of any taps, lubricated or otherwise, since lubricants would contaminate the surface and unlubricated taps either seize or else slip out under the weight of a large head of mercury.

---

* Bosworth's results are of course also explicable if we assume that only a fraction of the molecules striking the surface are adsorbed and that this fraction is different for different gases, but the available evidence does not appear to contradict the view that practically all the molecules which strike a metal surface condense thereon.

The flat chamber was prepared by grinding a circular depression in a slab of ⅜-in. plate glass and covering it with a similar slab, the two being sealed together by a special optical cement. After much trouble it was found that even such thick glass always yielded sufficiently under atmospheric pressure to crack the seals. Though leak could be prevented by an outer plastic coating round the seal there remained the probability of gas being entrapped in cracks in the seal and slowly escaping to vitiate the measurements, especially as the whole quantity of gas expected in the apparatus would be less than 0·001 c.c. at atmospheric pressure. Accordingly the flat chamber was replaced by a vessel made from a tube of pyrex glass of large size and the apparatus redesigned and constructed wholly of pyrex glass sealed into one piece. The new chamber only permitted a much smaller surface being formed (60 sq. cm.), but a monomolecular layer over this area would yield gas sufficient to give a pressure of about 0·0012 mm. in the apparatus and this could easily be measured with the required accuracy. Moreover the new apparatus could be strongly heated, thus making it much easier to remove water from the glass walls and so bring them to a steady state in regard to adsorption and emission of gases.

§ 2. APPARATUS AND METHOD

The vessel *B*, figure 1, was made from pyrex tubing 7 cm. in bore. It was fixed almost horizontally as shown, so that mercury from *A* could be allowed to flow into it, forming a surface of area 60 cm^2. The mercury reservoir *A* was situated about 3 ft. from *B* and connected thereto by glass tubes *C* and *D* about 2 mm. in bore. These were sufficiently flexible to permit *A* to be raised or lowered through a distance of 6 or 7 cm. The tube *D* allowed the gas in *A* to adjust itself to the same value as in the rest of the apparatus. The reservoir *A* was carried on the slide-rest of a lathe. This was firmly bolted to the bench in a vertical position, enabling *A* to be raised or lowered fairly quickly and without risk of exceeding the limit of flexibility of *C* and *D* or of imposing irregular strains on them. Thus the mercury could be caused to flow into or out of *B* without the use of taps or valves and independently of the gas pressure in the apparatus.

B was sealed directly to the Langmuir pump and McLeod gauge by wide tubing to allow rapid pumping and equalization of pressure in the apparatus. The pumps were not mounted on the same bench as the apparatus, so that their operation would entail a minimum of vibration of the mercury surface in *B*.

The narrow tubes *E* and *F*, open at the lower ends, dipped into vessels of mercury so that when the apparatus was exhausted the mercury stood just below the join as at *K*. Hence when the vessels *G* and *H* were raised, the chamber *B* and the gauge were isolated from the rest of the apparatus by mercury cut-offs, so that any gas evolved from the mercury could produce a maximum increase in pressure and be measured more rapidly. Even if a leak were to occur at the gas-inlet *M* or the mercury seal *L* it would not affect the reading when *G* and *H* had been raised. To attain rapid pumping it was necessary to avoid having to evacuate *A* through the long narrow tube *D* and this was done by raising *H* a few seconds after starting the pump, when the pressure would have fallen to a fraction of a millimetre.

During the construction of the apparatus the glass tubing was carefully cleaned, and the separate parts were again cleaned before being sealed together in situ. The whole was then steamed out and dried with an air current before being sealed to the pump. To remove occluded moisture from the walls the apparatus was heated for some time while subject to high vacuum. For this purpose *B* was surrounded by a woven electrical resistance and the narrow tubes were wound with thin wire and heated electrically. The wide tubes could be safely heated with a flame. The mercury was cleaned by distillation under reduced pressure in a slow air current and by washing, as described elsewhere(1, 8).

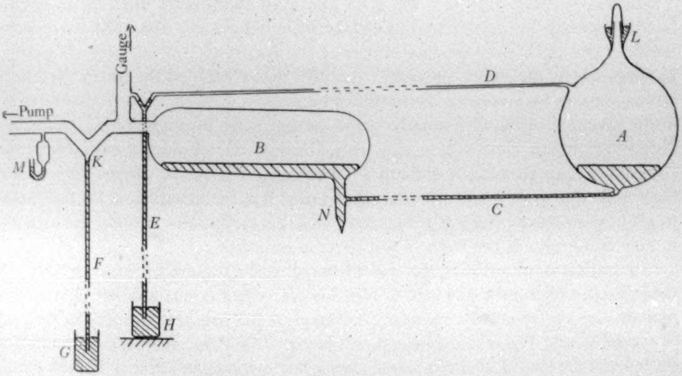


Figure 1.

To place the mercury in the apparatus, *A* was lowered as far as possible, some mercury was poured in, and the seal *L* was closed. The apparatus was then exhausted and the tube was made hot by means of its winding of wire, after which *A* was slowly raised. This avoided the possibility of any gas being trapped in *C* or *N*, the mercury being actually boiled in each of these. The mercury could be removed for re-distilling by snipping off the tip of *N*.

An experiment consisted essentially of the following operations: (*a*) gas was admitted up to the required pressure through inlet *M*; (*b*) *A* was raised and a mercury surface of area 60 cm^2 was thus formed in *B*; (*c*) after an interval the gas was pumped out, *E* and *F* were closed as has been explained, and the pressure was measured; (*d*) *A* was lowered causing the surface to collapse, the mercury flowing out till it stood at the top of the tube *N*; (*e*) the pressure of gas in the apparatus was measured.

§ 3. EFFECT OF ADSORPTION BY WALLS OF APPARATUS

In most previous work on the direct measurement of adsorption, the gas has been admitted at a low pressure only and the reduction of this pressure was noted as adsorption proceeded. The walls were not subjected to large changes of pressure and the area of adsorbing surface (e.g. in charcoal) was large compared with the walls, though this area was not usually directly measurable.

In the present experiment the mercury surface represents only a few per cent of the area of the inside of the apparatus. Moreover the walls are alternately subjected to atmospheric pressure of gases and then a pressure of 10^{-4} mm. Hence it was necessary to consider the possibility that adsorption and desorption of the glass might complicate or perhaps obscure the effects due to the mercury. Tests were made, without mercury in *B*, of the rate at which gas was evolved from the walls after these had been exposed to gas at varying pressures and for varying periods. These tests showed that once the glass had been thoroughly dried by prolonged heating in a high vacuum it became consistent in behaviour. Since the experiments involved the use of only dry gas, and since the apparatus did not vary appreciably in temperature, the simple procedure outlined below enabled the gas evolved from the mercury to be measured with sufficient accuracy. Observations were also made of the effects of raising the temperature of the apparatus through 100° C., and it was found that though consistent results were obtained the amount of gas evolved was sufficiently large to make it difficult to disentangle the effects due to the mercury from those due to the walls. An interval of 5 min. was found sufficient for the pumps to give the necessary stage of exhaustion, and this period also allowed the mercury to flow either into or out from *B* and become stationary.

In making an observation gas was admitted to the required pressure and *A* was then raised, causing the mercury to flow into *B*, where it was allowed to stand for periods varying from a few minutes (the time for the mercury to come to rest) up to several hours. The condensation pump being already hot, the backing pump was started and the cut-off at *E* was raised after a few seconds and that at *F* after 5 min. The pressure was then read on the gauge and the reading was repeated, usually at 5-min. intervals. During one of these intervals *A* was lowered, destroying the mercury surface in *B*. The graph of pressure against time shows by a sudden rise in this interval the effect of the gas liberated from the mercury, and this can be quite well separated from the slow evolution of gas from the walls, even though this factor varied considerably with the pressure of gas and the time it had been retained in the apparatus. By variation of the period between exhaustion and collapsing of the surface it was shown that the adsorbed gas was retained by the mercury for a considerable time.

§ 4. ADSORPTION AND LIFE OF CARBON DIOXIDE ON THE SURFACE OF MERCURY

The gas was admitted from a high-pressure cylinder through a series of drying tubes and filters of glass wool. A thin rubber diaphragm over a side tube acted as a safety valve to protect the apparatus against high pressures.

A pressure of 10^{-4} mm. would correspond to a total of 29×10^{14} molecules at laboratory temperature in the whole volume of *B*, the gauge, and the wide connecting tube. A monomolecular layer of gas over the 60 cm.^{2} of mercury would, on complete desorption, yield approximately 34×10^{15} molecules and thus cause a rise in pressure of 12×10^{-4} mm.

In figure 2 a typical selection of curves is given showing the rise in pressure which occurred on collapse of the surface, the time in each case being measured from the instant when the cut-off *F* was raised. Other details of the experiments are given below the figure. The curve *B* indicates the small amount of adsorption detected when the mercury surface was formed in a fairly high vacuum and the carbon dioxide subsequently admitted. A comparison of the curves shows how much less readily adsorption occurs under such conditions than in the case when the surface is formed in the presence of the gas. This fact is in keeping with comments by writers on the differences observed between a static mercury surface and one that is expanding or freshly formed. The observed rise in pressure due to liberation of carbon dioxide varied from about 4×10^{-4} mm., when the surface had been formed

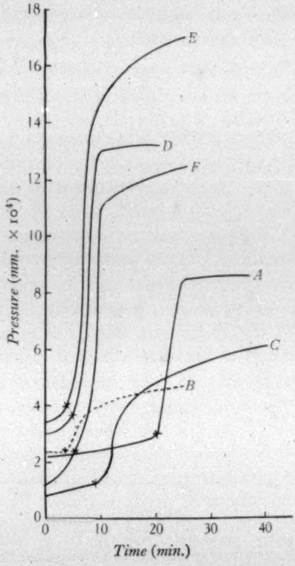


Figure 2. × indicates point at which surface was collapsed. *A*: CO_2, $p = 10$ cm., exposure 2 h. *B*: Hg admitted first, then 12 cm. of CO_2. *C*: CO_2, $p = 8$ cm., exposure 48 h. *D*: CO_2, $p = 36$ cm., exposure 20 min. *E*: CO_2, $p = 76$ cm., exposure $2\frac{1}{2}$ h. *F*: CO_2, $p = 76$ cm., exposure less than 10 min.

in carbon dioxide at a pressure of 8 cm., to almost 12×10^{-4} mm. when the surface had been formed in the same gas at atmospheric pressure. The latter amount corresponds to practically a complete monomolecular layer. The time of exposure, as was expected, had little effect on the amount of gas retained by the surface. It is just possible that very long exposure resulted in less gas being recovered (as in *C*) owing perhaps to something replacing the carbon dioxide on the surface. The observed variation of the amount adsorbed with the pressure of gas is hardly to be expected, however, if the adsorbed layer of gas is retained by the surface; see § 6. On a number of occasions the mercury surface was formed in gases at pressures of about 0·001 mm., but in no case did measurable adsorption occur at this pressure under the experimental conditions pertaining.

§ 5. MEASUREMENTS ON HYDROGEN AND DRY AIR

Experiments were also carried out with hydrogen, which was passed through a tube containing heated platinized asbestos before reaching the drying train. Measurements showed the same general features as in the case of carbon dioxide; the gas being retained by the surface on evacuation as was indicated by the observations of Cook⁽¹³⁾ though there was some suggestion that the mercury surface would lose some of the adsorbed hydrogen more easily than it lost carbon dioxide. Curves *G* and *H*, figure 3, are for readings made some months apart. The curve *K* is for air admitted through the drying tubes and indicates an adsorption of the same order as for hydrogen, though it is of course possible that the adsorbed gas did not have the same proportions of oxygen and nitrogen as the air. Curve *L* is for carbon dioxide when a period of nearly half an hour elapsed between the evacuation and collapse of the surface. The flatness of the curve both before and after collapse of the surface indicates that the adsorbed layer was not being lost at an appreciable rate in a vacuum of 2×10^{-4} mm., and also shows that very little gas was evolved from the glass, at any rate after the first 5 min.

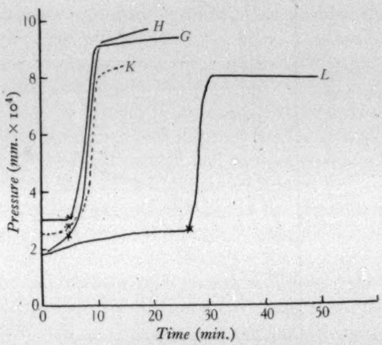


Figure 3. *G*: H_2, $p=40$ cm., exposure 35 min. *H*: H_2, $p=40$ cm., exposure 35 min. *K*: Air, $p=45$ cm., exposure 5 min. *L*: CO_2, $p=50$ cm., exposure 10 min.

Tests were made at times of the effect of bumping the apparatus before collapsing the surface and it was found that the surface could be caused to oscillate a good deal without any appreciable evolution of gas.

§ 6. DISCUSSION OF RESULTS

When a mercury surface is formed in any of the gases carbon dioxide, hydrogen, or air an adsorption occurs at the surface, the limit apparently being a monomolecular layer, which in these experiments was only approached as the gas approached atmospheric pressure. Little, if any, other work has been published with reference to direct measurements of adsorption of gases on mercury. Both Oliphant and Bosworth measured the amount of gas adsorbed selectively from mixtures at atmospheric pressure, and by using a shower of drops ensured that they were always dealing with a freshly formed surface. This is an important consideration since even under the very best experimental conditions the properties (e.g. the surface tension and the photo-electric effect(20)) of a mercury surface at rest are seldom found to agree with those of an expanding surface.

Sissingh and Haak(11) in 1919 observed the change in optical properties of mercury with time of exposure to air and estimated that the thickness of the adsorbed layer was about $1{\cdot}6\mu\mu$. More recently E. Herschkowitch(12), working on the same subject, has reported that neon and nitrogen produce no measurable effect but that hydrogen, oxygen, and carbon dioxide appear to be absorbed in layers varying in thickness from 5 to 15 molecular diameters. The present work makes it clear that if such thick layers do condense on the surface then all layers beyond the first must be so loosely bound as to be at once desorbed on evacuation, whereas the first layer appears to have an almost indefinite life on the surface. Herschkowitch(12) does in fact record a partial recovery of the mercury surface when the space above it is evacuated, which suggests that part of the observed optical effect is due to an adsorption which is reversed on evacuation.

Stoeckle(14) records an evolution of gas from a mercury surface and Cook(13) records the following observations in his paper on the surface tension of mercury: "In one case the tension of mercury in contact with hydrogen at 0·00015 mm. remained constant at 421 dynes over a period of 8 min. On spilling a drop from the mercury (thus breaking the surface) the tension immediately rose to 463 dynes and remained at this value for 14 min. Simultaneously the pressure of hydrogen rose to 0·00048 mm.;" and later "the tension in hydrogen decreased from 475 dynes to 484 dynes in 20 min. while the pressure fell from 0·0037 mm. to 0·0020 mm." The significance of the periods of time stated is not clear, but it does appear that under the conditions which must be attained for accurate surface tension measurements the mercury adsorbs hydrogen at quite low pressures. The figures given by Cook show that an evolution of gas sufficient to raise the pressure by 0·00033 mm. was accompanied by a rise of 52 dyne/cm. in surface tension, yet a fall of only 11 dyne/cm. was accompanied by a fall in pressure 0·00170 mm. If these figures could be taken as due to effects occurring at the mercury surface only, it would mean that the adsorption per fall of 1 dyne/cm. in surface tension was more than 25 times as great as the amount desorbed for an equal rise in surface tension. In the absence of any other information about the actual conditions obtaining, it seems most probable that all the gas evolved came from the mercury but that the glass walls were contributing to the observed adsorption, since it is unlikely that actual solution of the gas in the mercury could account for more than a fraction of the difference.

The significance of the variations of the amount of gas evolved from the surface with the pressure of gas to which the surface was exposed is by no means clear. Since the adsorbed molecules have a long life on the surface, it is to be expected that a complete monomolecular layer would be adsorbed and retained from gas at a pressure of 8 cm. equally as well as from gas at atmospheric pressure. There is, however, a striking parallelism between these results and those obtained for the surface tension by some of the workers using the method of the sessile drop. Thus Popesco(15) found the initial value of surface tension highest for a surface formed in gas at atmospheric pressure, with lower values when the surface was formed in gas at lower pressures, the lowest value being for the surface formed in a vacuum; but that the subsequent admission of gas to the surface did not cause any rise in the surface tension. Consistently with this the author finds that very slight adsorption occurs when the surface is formed in a vacuum and the gas admitted later, and that for a surface formed in the gas the adsorption is larger for higher pressures.

Explanations of the above-mentioned variations of surface tension have been advanced in terms of the orientation of the molecules to positions of minimum energy. These explanations are open to objection and it is in fact now found that the surface tension of mercury is not lower in a vacuum than in gases. In spite of the work of Hiss and others, N. K. Adam(16) concludes that there is no satisfactory evidence that the surface tension of any pure liquid changes with time. The low values of surface tension of mercury must be ascribed to some factor acting as a contamination, though it has never been possible to establish the nature of this contamination. In the present work, owing to the effect of the glass it was useless to proceed to the very high vacuum obtained by prolonged degassing, and it is probable that the same influences which caused the lower surface tensions at reduced pressures were present here, this contamination occupying some of the surface and preventing adsorption at very low pressures. Cook, was able, however, to detect the adsorption when working under the conditions necessary for surface-tension determinations. The present work shows that it is possible to measure adsorption on mercury by a direct method. The advantage of having a perfectly definite area of surface (as contrasted with the surface-area of charcoal, powders, or crystalline aggregates) is largely offset by the uncertainty which still exists as to the actual cause of the variations to which reference has just been made. Evidence indicates, however, that adsorption approximately to the extent of a complete monomolecular layer occurs on mercury surfaces formed in gases at moderate pressures, the adsorption probably being complete by the time the surface has come to rest. This layer persists on evacuation in spite of considerable oscillation of the surface, but is liberated when the surface is destroyed.

Freundlich accepts the view that the fall in surface tension in gases is due to adsorption, and says that no fall of value with time occurs in a vacuum, as is to be expected with an unassociated liquid. No suggestion is offered as to the cause of the higher initial values for gases which are quoted. Freundlich further states that the adsorption is very rapid, "the process requiring a finite time being a diffusion in a very thin layer." A critical survey of all the available data on the effect of gases on the surface tension on mercury shows that agreement between workers is far from satisfactory. There seems no doubt that a monomolecular layer is adsorbed practically before any appreciable fall in surface tension has occurred. If this single layer is the cause of the fall in surface tension observed by various workers over periods of hours or days, the adsorption and surface tension can scarcely be connected by an equation such as that of Gibbs.

On the other hand if the subsequent fall in surface tension is due to the adsorption of polymolecular layers (the existence of which is suggested by Herschkowitch and others) it is not to be expected that these layers should be desorbed so readily in comparison with the initial layer. It is of course possible for a large reduction of surface tension to be caused by a layer that is lost on evacuation. Some unpublished observations by Prof. Kerr Grant show that the surface tension of water falls by over 20 per cent. in the presence of ether vapour, but recovers almost completely as soon as the ether vapour is allowed to escape. Moreover S. G. Cook does record a considerable recovery in the surface tension of mercury which has been exposed to air or oxygen if the pressure of the gas is lowered, though it never rises to the initial value for a vacuum. The mercury requires a long time to reach equilibrium with a given pressure of gas whether the pressure is being increased or decreased. These observations of Cook suggest a slow adsorption that is more or less reversible as a possible cause of the variation in surface tension, though any such slow desorption was not detected in the present work.

§ 7. CONCLUSIONS

The conclusions reached in this paper may be summarized as follows. Adsorbed carbon dioxide or hydrogen to the extent of a monomolecular layer is retained by the mercury surface on evacuation. The adsorption is readily observed when the mercury surface is formed in a gas but not when the gas is admitted to a surface which has been formed in a vacuum. It seems more probable that this is due to something acting as a contamination than to some inherent property of the mercury causing changes in the transition layer at the surface. If more than a single layer of gas molecules condenses on the surface the subsequent layers are lost on evacuation.

Though the mercury surface, from its uniformity and definite area, should be ideal for quantitative work on gas-adsorption on a metal, the phenomena observed are at present too complex to be embraced in a single theory. For processes occurring during the actual formation of the surface it is probable that the theory of adsorption at liquid surfaces applies, but this sheds little light on processes occurring after the surface has come to rest.

§ 8. ACKNOWLEDGMENTS

The author's thanks are again due to Prof. Kerr Grant for helpful discussion and to Mr R. H. Oliphant for his skill in constructing the apparatus in such a way that it could withstand the severe strains imposed on it.

REFERENCES

(1) Burdon. *Proc. phys. Soc.* **38**, 148 (1926).
(2) Burdon and Oliphant. *Trans. Faraday Soc.* **23**, 205 (1927).
(3) Oliphant and Burdon. *Nature*, Lond., **120**, 584 (1927).
(4) Mitton. *Trans. roy. Soc. S. Afr.* p. 267 (1929); Gibson, p. 51 (1932).
(5) Oliphant. *Phil. Mag.* **6**, 422 (1928).
(6) Burdon. *Nature*, Lond., **129**, 456 (1931).
(7) Bosworth. *Trans. Faraday Soc.* **28**, 896 (1932).
(8) Burdon. *Trans. Faraday Soc.* **28**, 866 (1932).
(9) Harkins and Ewing. *J. Amer. chem. Soc.* p. 2539 (Dec. 1920).
(10) Iredale. *Phil. Mag.* **45**, 1088 (1923).
(11) Sissingh and Haak. *Proc. Acad. Sci. Amst.* **21**, 678 (1919).
(12) Herschkowitch. *Ann. Phys.*, Lpz., p. 993 (Aug. 1931).
(13) Cook. *Phys. Rev.* **34**, 513 (1929).
(14) Stoeckle. *Ann. Phys.*, Lpz., **66**, 499 (1898).
(15) Popesco. *Ann. Phys.*, Paris, **3**, 402 (1925).
(16) Adam. *Physics and Chemistry of Surfaces*, p. 152.
(17) Freundlich. *Coll. and Cap. Chem.* p. 82.
(18) Tronstad and Feachem. *Proc. roy. Soc.* **145**, 115 (1934).
(19) Tronstad. *Trans. Faraday Soc.* **29**, 502 (1933).
(20) Dunn. *Phys. Rev.* **29**, 639 (1927).

621.3.018.362 1.396.645.22

A NEGATIVE-RESISTANCE DEVICE AND ITS APPLICATION TO HARMONIC ANALYSIS

By C. W. OATLEY, M.A., M.Sc.
University of London, King's College

Received November 24, 1934. *Read March* 1, 1935.

ABSTRACT. A description is given of a negative-resistance device which is very similar in principle to the kallirotron amplifier described by L. B. Turner. It is shown that when this device is used to decrease the effective resistance of a tuned circuit no appreciable distortion of the resonance curve of the circuit is produced. A method is given whereby the effective resistance of the tuned circuit may quickly be adjusted to any predetermined value.

§1. GENERAL PRINCIPLES

Negative-resistance devices have found many interesting applications since the introduction of the dynatron by Hull in 1918. Shortly after this date Turner published an account* of a circuit, termed by him the "kallirotron amplifier", in which a negative-resistance effect was obtained without the aid of secondary electron emission. About five years ago the present author, who at that time had not seen Turner's paper, devised a very similar circuit which has since been used in the manner described below. In view of the facts that Turner's paper is not very generally accessible and that the kallirotron differs in certain practical details from the author's circuit, a brief description of the latter will first be given.

Consider the circuit of figure 1. The condensers C_1 and C_2 are inserted merely for the purpose of preventing the flow of direct current and we suppose them to be so large that their reactances are negligible. The resistances R_1, R_2 and R_3 are so R_1, R_2, R_3
large that currents flowing through them may be neglected in comparison with the anode current of the valve V_2. Suppose an alternating e.m.f. of magnitude E to E
be applied between the terminals A and B. Then some fraction bE of the e.m.f. is b
applied between grid and filament of V_1, where b is a constant lying between 0 and 1 and depending on the setting of the potentiometer R_1. Since R_2 is large, the e.m.f. applied between the grid and filament of V_2 will be nearly equal to $-\mu bE$, where μ is the amplification factor of V_1. If the mutual conductance of V_2 be k the μ, k
alternating component of current flowing through this valve will be $-\mu kbE$. Since currents in other branches of the circuit are negligible in comparison with this, we have the result that the application of the e.m.f. E gives rise to a total current $-\mu kbE$. Thus, looked at from the terminals AB, the circuit is equivalent to a resistance $-1/\mu kb$ so far as alternating currents are concerned. Taking as typical

---

* L. B. Turner, *Radio Review*, 1, 317 (1920).

values $\mu = 25$, $k = 1$ mA./V., and remembering that b can vary between 0 and 1, we see that the value of the negative resistance can be made to lie anywhere between -40 and $-\infty$ Ω.

Comparing the present arrangement with a dynatron we see that both require the circuit in which they are connected to offer a passage to direct current. With regard to frequency-range the present device will operate successfully from about 20 c./sec. to 100 kc./sec. At higher frequencies inter-electrode capacities become important; with care the upper limit might be extended, but experiments in this direction have not been carried out. Below 20 cycles per second the blocking condensers have to be made inconveniently large, but it is possible to use battery coupling instead. On the whole, however, the dynatron is to be preferred if the range of frequencies to be covered is very wide. With regard to stability of characteristics, the present device has the great advantage that, unlike the dynatron, it does not depend upon the somewhat erratic phenomenon of secondary emission, and there is no reason why it should vary in performance any more than does a simple low-frequency amplifier.

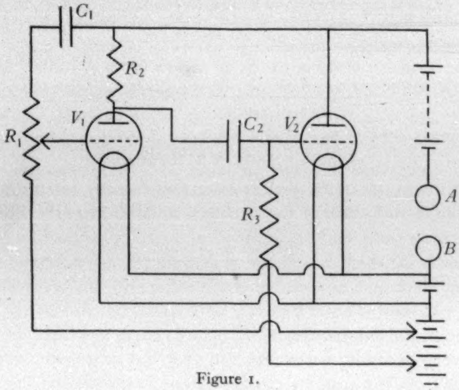


Figure 1.

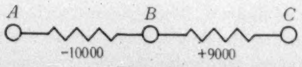


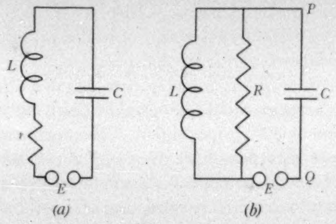
Figure 2.

Another very important advantage over the dynatron lies in the range of values of negative resistance which can be obtained. With ordinary receiving valves it is difficult to obtain by the dynatron method resistances lower than some 10000 Ω. in absolute value, whereas with the present device negative resistances as low as -40 Ω. are easily possible. Admittedly one can always obtain a low value of negative resistance by placing a positive resistance in series with a dynatron, but the practice is not to be recommended. Thus in figure 2 if we place a positive resistance of 9000 Ω. in series with a negative resistance of -10000 due to a dynatron, the effective resistance between A and C is -1000 Ω. However, if the maximum permissible voltage across the dynatron be V, the voltage between A and C must not exceed $V/10$ and this limitation may be serious. Furthermore, if the resistance of the dynatron should change accidentally by 1 per cent, the resistance between A and C would change by 10 per cent.

One final advantage of the present device is the ease with which the magnitude of the negative resistance can be changed merely by varying the setting of the potentiometer R_1.

§ 2. APPLICATION TO HARMONIC ANALYSIS

The simplest type of electrical harmonic analyser is that in which the various components of a complex current wave are selected one by one by means of a sharply tuned resonant circuit and measured with a thermoammeter or other suitable device. In practice the chief difficulty is to obtain circuits of sufficiently low resistance to enable the various components of current to be effectively separated. However, this difficulty can be overcome by the use of the negative-resistance device described above.



(a) (b)

Figure 3.

To a first approximation, the resonant circuit of figure 3 (*a*) may be replaced by the equivalent circuit of figure 3 (*b*), where $R = L^2\omega^2/r$ and the approximation will be sufficiently good for our present purpose provided that $L\omega/r$ be not less than 10. Coils for which this is the case are not difficult to construct except at the lowest acoustic frequencies. If now we connect the points P and Q, figure 3 (*b*), to the terminals A and B of the negative resistance circuit of figure 1, we shall place a negative resistance, R' say, in parallel with R. The resulting equivalent shunt resistance of the tuned circuit will be $RR'/(R+R')$, and theoretically this can be made as large as we please by suitable choice of R'. The practical limit is set by the tendency of the circuit to oscillate spontaneously.

L, ω, r

The above arrangement is, of course, simply one method of applying reaction to a tuned circuit. Its advantage over most other methods lies in the ease and accuracy with which the reaction can be controlled, the freedom from any distuning effect due to reaction, and in the fact that it can be used with circuits tuned to quite low frequencies.

§ 3. PRACTICAL DETAILS

If instability is to be avoided, the useful range of values of negative resistance will be from $-R$ to $-\infty$, and in most cases R will not be less than about 10000 Ω. If the circuit of figure 1 be used as it stands, the whole useful range of variation of the negative resistance will be covered by a very small movement of the potentiometer slide, and fine control will be impossible. The minimum obtainable value of negative resistance could be increased to 10000 Ω. by reducing the values of R_1, R_2, or R_3, but such methods are objectionable because they amount to placing a positive resistance in parallel with the negative one. In such a case a small accidental variation of either the positive or the negative resistance would cause a much larger percentage variation of the combination of the two. A better method of increasing the minimum value of negative resistance is to place a suitable resistance in the anode circuit of V_2 and another between R_1 and C_1. If the first be made several times as large as the slope resistance of V_2 it will greatly reduce the effect of any accidental variation of this quantity (due, for example, to fluctuation of battery voltages), while the second resistance in series with R_1 can be chosen so that the requisite range of values of negative resistance is just covered by the full movement of the potentiometer slide.

§ 4. EXPERIMENTAL RESULTS

Since a harmonic analyser should be capable of giving quantitative results, it is necessary to ascertain whether the addition of the negative-resistance device to a tuned circuit causes any distortion of the resonance curve apart, of course, from the change to be expected from a decrease in the effective resistance of the circuit. To test this point a series resonant circuit consisting of a coil having an inductance of 1 H. and a condenser of capacity 0·0397 μF. was set up. A small e.m.f. of the correct frequency for resonance, about 800 c./sec., was induced in the coil and the p.d. across the condenser was measured with a valve voltmeter. The negative-resistance device, connected in parallel with the coil, was adjusted until the circuit was sharply tuned. A variable air condenser connected across the main condenser enabled the tuning of the circuit to be varied over a small range of values, and in this way data for a resonance curve were obtained. The results are shown in the table, where the figures in the first column refer to settings of the air condenser and those in the second to the readings of the valve voltmeter. The values in the last column are calculated from the formula for a series-tuned circuit. In applying this formula we need to know the effective resistance of the circuit and also the e.m.f. induced in the coil. These two quantities are calculated from the observed maximum and minimum readings of the valve-voltmeter, marked with an asterisk in the table. It will be seen that the remaining voltmeter readings are in good agreement with the calculated values. In this particular example the effective resistance of the circuit is found to be about 6·7 Ω., corresponding to a voltage-magnification at resonance of 750, which is many times as great as would have been obtained without the use of the negative resistance.

θ	V (observed)	V (calculated)
69	40·0*	40·0
68	48·0	48·7
67	61·0	59·1
66	76·0	73·4
65	91·0	88·3
64	103·0	101·5
63	110·0*	110·0
62	108·0	108·0
61	98·0	101·0
60	85·0	87·7
59	72·0	72·7
58	57·0	58·5
57	48·0	48·7
56	40·0	39·8

§ 5. DETERMINATION OF THE EFFECTIVE RESISTANCE OF THE CIRCUIT

When a series-tuned circuit is used in conjunction with the negative-resistance device, the effective circuit resistance can always be found by the method described above, but this would be too tedious when the circuit was being used to separate the various components of a complex voltage wave, since the effective resistance would alter every time the circuit was tuned to a fresh component. The following more direct method was therefore tried.

If the negative resistance be gradually decreased, a point will be reached where the circuit is on the verge of spontaneous oscillation. In practice this point is quite clearly defined and obviously corresponds to an effective circuit resistance of zero. If, having brought the circuit to this state, we now insert an extra resistance P in series with the inductance, then P should be the total effective circuit resistance and can be chosen to have any desired value. Preliminary tests of this method gave unsatisfactory results and the trouble was eventually traced to the fact that the resistance P is traversed by the direct current flowing in the negative resistance device and the addition of P therefore changes the value of the negative resistance. The difficulty was overcome by placing P in series with the condenser instead of with the inductance.

As a quantitative test of this method the following experiment was carried out. The tuned circuit (inductance 1 H., capacity 0·1 μF.) was brought to the verge of oscillation by adjustment of the negative resistance and a resistance of 1 Ω. was then connected in series with the condenser. An e.m.f. of the resonant frequency was next induced in the coil and adjusted in magnitude till the valve voltmeter showed a full-scale reading. The resistance in series with the condenser was then increased to 2 Ω. and, as a result, the voltmeter reading dropped to 0·51 of its previous value. The experiment was then repeated, the resistance in series with the condenser being at first 2 Ω. and subsequently increased to 4 Ω. This time the first voltmeter reading was exactly twice the second. Two further steps gave the same result and brought the nominal effective circuit resistance up to 16 Ω. The actual effective circuit resistance was then determined by the distuning method and found to be 15·6 Ω. The difference between these two values is probably not greater than the experimental error in the second determination. It thus appears that the method described above provides a satisfactory means of adjusting the effective circuit resistance to any desired value.

DISCUSSION

Dr L. Hartshorn asked whether the same circuit would serve for the whole range of audio frequencies.

Dr A. B. Wood asked whether the negative resistor contained any reactive component; if so, it might make a good oscillator.

The Author replied that the apparatus worked satisfactorily over the range from 100 to 10,000 c./sec. The negative resistor contained only a negligible reactive component if the condensers C_1, C_2 were large enough, but could be converted into a good oscillator for the range from 50 to 10^5 c./sec.

621.383.5

A NEW SELENIUM-SULPHUR RECTIFIER PHOTOELECTRIC CELL

By G. P. BARNARD, B.Sc., A.Inst.P., Grad.I.E.E.
The National Physical Laboratory

ABSTRACT. The paper describes a method by which a new selenium-sulphur rectifier photoelectric cell may be constructed, and gives the results of the first part of a full investigation of the physical behaviour of these new cells. The results are related to the details of construction of the cell, so as to show how far they support the theories of A. H. Wilson and of Frenkel and Joffé. A number of simple equations expressing the performance of the cell have been deduced.

---

§ 1. INTRODUCTION

The comparatively recent appearance of the rectifier or *Sperrschicht* photoelectric cell has recalled into prominence the contact photoelectric phenomena observed by Uljanin.(1) Uljanin annealed a thin layer of selenium between two plates of glass; the glass surfaces in contact with the selenium were covered with semitransparent metal films which formed the electrodes of the cell. He found that when the selenium was illuminated through one of the semitransparent metal films the cell generated an electromotive force; with sunlight the open-circuit e.m.f. generated amounted to 0·12 V. The illuminated electrode was the negative pole of the cell.

This cell should not be confused with the well-known photoconducting selenium cell, which shows an increase of conductance on illumination. The photoconducting cell does not generate an e.m.f. on illumination and is always used with an external source of e.m.f.

The modern Sperrschicht cell is said to consist of a metal plate coated with a thin layer of some semiconductor, the upper surface of which is covered with a very thin film of metal. The metal plate and the metal film form the electrodes of the cell and are directly connected to the terminals of a current-measuring instrument. Precise details of construction do not appear to have been published either in scientific journals or in patent specifications. Such information as is available is not sufficient for the construction of a cell with a sensitivity comparable with the known sensitivities of cells on the market.

The work described in the present paper was originally undertaken in order to develop a simple and accurate instrument which could be used for measuring the distribution of illumination in small-scale model rooms. The original intention was to use one of the well-known forms of Sperrschicht cell; but it was immediately apparent that these would be unsatisfactory for the measurement of diffused light, owing to their marked decrease of sensitivity with increase of angle of incidence.

Accordingly it was decided to investigate the method of construction of such cells in order, if possible, to effect an improvement in this respect. A simple method was eventually devised which enabled compensation for this decrease in sensitivity to be made to within 3 per cent of the required sensitivity for angles of incidence from 0 to 85 degrees; the results of this work will be the subject of a separate publication. The present paper gives the results of the first part of a full investigation of these new cells.

In a recent short note⁽²⁾ the author has described briefly a method of constructing a new selenium-sulphur rectifier photoelectric cell. The description is now given for the first time in sufficient detail, and the construction of the cell is related to the results obtained, so as to show how far these results support the theories of A. H. Wilson, and of Frenkel and Joffé. A number of simple equations expressing the performance of the cell have been deduced.

§ 2. MODERN THEORY

A short qualitative discussion of the theoretical treatment given by certain writers of the rectifying properties of a semiconductor/metal contact will serve to indicate the general nature of the problem, and to prepare the way for the interpretation of the experimental results described in this paper. A list of original papers is given for reference⁽³⁾.

2·1. *The contact gap.* It is always assumed at the outset that in the system metal/semiconductor/metal the two contacts are dissimilar; that at one contact the bounding surfaces of the metal and the semiconductor are separated by a small gap δ of definite width δ, while at the other contact the separation of the bounding surfaces is relatively so small that this contact can have none of the properties of the first. Thus with this assumption it is only necessary to consider the first of these contacts.

It is not possible to say definitely what is the nature of the surface of the semiconductor that in contact with a metal will yield a high-resistance interface; experimentally, however, it is known that certain methods of treatment give high-resistance contacts, and that other methods do not. Schottky and Deutschmann's measurement of the electrostatic capacity of the transition layer showed that δ, in the cells used by them, was of the order of the mean free path of electrons in ordinary electronic conductors. But little or nothing is known of the actual nature of this transition layer. That it is not a vacuum, or a layer of gas, is admitted generally, although in its properties it can be regarded as being mathematically equivalent to a vacuum. N. R. Campbell and A. H. Wilson adhere to the view that the bounding surfaces of the metal and the semiconductor are separated by an insulating film of the semiconductor freed from all conducting impurities. So long as the thickness δ of this film is smaller than the mean free path of an electron in the material of the film, the film will be exactly equivalent to a layer of vacuum with a reduction, depending upon the dielectric constant of this insulating film, of the height of the potential-energy mountain, which results from the image-force effect, between the surfaces of the semiconductor and the metal.

2·2. *Mathematical model of semiconductor*. According to A. H. Wilson's treatment, all electrons in the semiconductor at absolute zero of temperature are in what may be called bound states. At any other temperature a few electrons are thermally excited into energy states in which they can conduct. The mean value of the potential energy of these electrons, regarded as nearly free, is represented by the energy level W_2, figure 1. Wilson then deals with the case of a lattice possessing atoms of an impurity, each atom having a single electron in a discrete state of energy W_1. The level W_1 lies below the level W_2 but above the highest occupied energy level, W, of electrons possessed by atoms of the pure semiconductor, so that electrons from the level W_1 can make the transition to the level W_2 more easily than electrons from the level W. The conduction electrons are thus entirely derived from impurities. His equations hold so long as the number of impurities is less than 10^{17} per unit volume but still sufficiently large to mask entirely the natural conductivity of the semiconductor. With these premises A. H. Wilson determines the distribution function for the impurity electrons.

W_2
W_1
W

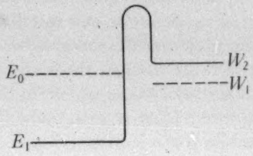


Figure 1. Potential-energy states at boundary between metal and semiconductor (Wilson).

Frenkel and Joffé on the other hand make use of the Maxwellian form of the Fermi-Dirac statistical expression for a small free-electron concentration in the semiconductor. They find, however, that for any reasonable value of the free-electron concentration, of the order of 10^{17} per cm³, the free energy of an electron turns out to be negative and at 300° K. is of the order of $-0·15$ electron-volt. The final equations deduced by Frenkel and Joffé and by Wilson to express the rectifying properties of a semiconductor/metal contact are not essentially different in so far as their application to any practical case is concerned.

2·3. *The transmission coefficient of the contact gap*. When the steady state has been reached for the flow of electrons in any given direction through and normal to the contact, the charge-distribution and the current may be obtained from a solution of the Schrödinger wave equation. The solution will in general show two terms, one representing an electron wave travelling in the positive direction—the incident wave—and the other representing a wave proceeding in the negative direction—the reflected wave. The ratio of the intensity of the reflected wave to that of the incident wave is the reflection coefficient R of the contact gap; while the ratio of the intensity of the transmitted wave to that of the incident wave is the transmission coefficient D. Obviously $D+R=1$. It is seen, therefore, that the

R
D

contact resistance results from the reflection of electrons at the assumed transition layer between the bounding surfaces of the metal and the semiconductor; on the other hand the bulk resistance of a solid arises from the interaction of the electrons with the elastic vibrations of the lattices. The transmission coefficient, and hence the resistance of the contact gap, depends upon the applied electric field in the gap.

2·4. *Equations to express the rectifying properties of contact.* In both cases the problem has been made one-dimensional, doubtless to simplify the mathematical treatment; consequently in the formulation of the results for any actual cell some modification of the equations so derived will be necessary. Figure 1 gives a one-dimensional picture of the potential states at the boundary between a semiconductor and a metal. The continuous curve represents the potential energy of conduction electrons, the left- and right-hand horizontal portions corresponding to the potential energies in metal and semiconductor respectively. Sommerfeld and Peirl's model of a metal is used, the valency electrons being considered as nearly free. The mean E_1 value of the potential energy of a valency electron is represented by E_1; its position relative to W_2, defined in 2·2, is determined from the condition of equilibrium. The level E_0 in the metal corresponds to the maximum critical energy of the Fermi distribution; E_0 is approximately equal to the maximum kinetic energy of an electron at absolute zero of temperature, and for most metals is of the order of 10 electron-volts. It is not necessary for the energy of an electron to be raised above the surface-potential jump in order to cross the surface into the transition layer, represented in figure 1 by a potential hump which, in the theoretical treatment, is assumed to be of rectangular form. Electrons can, according to the wave-mechanical treatment, pass through the hump if it is not too high or too broad. An analogous case is the passage of light-waves from glass into air and into another piece of glass, even at angles greater than that of total internal reflection, when the intervening air film is sufficiently thin.

Electrons passing from the metal to the semiconductor must acquire kinetic energies of not less than W_2; electrons passing from the semiconductor to the metal must also be raised by thermal excitation to the level W_2, in order to become conduction electrons in the semiconductor. If there is no potential difference applied between the metal and the semiconductor they will assume a contact difference of potential in the equilibrium condition for zero field, when equal numbers of electrons flow in either direction. The equations of A. H. Wilson show that the current from the semiconductor to the metal is equal to the current from the metal to the semiconductor only when E_0 lies approximately half-way between W_1 and W_2. This is, then, the equilibrium condition for zero field, the position of E_1 with respect to W_2 being determined.

V If now the potential of the metal be raised by a small quantity eV less than $(W_2 - E_0)$, the effect of the applied field will be to change the transmission coefficient and to increase E_1 and E_0 by eV, while leaving W_2 unaltered. It is no restriction to consider W_2 as invariable, since differences only in these quantities are of significance. The raising of E_0 above the equilibrium position for zero field will be compensated by an increased flow of electrons from the metal to the semiconductor, while leaving unchanged the current in the reverse direction. The difference between these two currents gives the net current from the metal to the semiconductor. If the subscripts a and b denote the semiconductor and the metal respectively, this net current ${}_bI_a$ is given by

a, b ${}_bI_a$

$${}_bI_a = \alpha \{\exp(-\beta V)\}\{\exp(eV/kT) - 1\} \qquad \text{......(1)},$$

where α is a constant depending upon T and upon the magnitude of the contact resistance at zero applied voltage, e the electronic charge, k Boltzmann's constant, T the absolute temperature, and β a constant determining the dependence of the transmission coefficient on V.

α e, k T, β

If the potential of the metal is lowered by the same quantity V, the net current from the metal to the semiconductor is similarly determined and is found to be negative. The current equation now becomes

$${}_aI_b = \alpha \{\exp(\beta V)\}\{1 - \exp(-eV/kT)\} \qquad \text{......(2)}$$

for the reverse direction.

Frenkel has deduced a value $\beta = 1{\cdot}5$.

It can be seen from equations (1) and (2) that the effect of the applied field on the transmission coefficient of the gap always tends to nullify the rectifying effect; this follows from the fact that a decrease in $(W_2 - E_1)$ makes the potential hump higher and decreases the transmission coefficient.

2·5. These equations may now be applied to the case where light falls upon a rectifier cell in the absence of any applied voltage. The metal in contact with the semiconductor is now sufficiently thin to transmit visible radiation. A number of electrons per second, extracted by light of energy F from the semiconductor, will cross the transition layer and enter the metal, this number being directly proportional to F. This electron current, from semiconductor to metal, will increase the contact potential difference between a and b by an amount V', which in its turn will produce a current in the reverse direction. If there is no other connexion between a and b, V' will reach a stationary value in the equilibrium condition when the electron-flows in the two directions are equal. If i_p denotes the primary photoelectric current and i_r the reverse current, equation (1) gives

F V' i_p i_r

$$i_p = i_r = \alpha \{\exp(-\beta V')\}\{\exp(eV'/kT) - 1\} \qquad \text{......(3)}.$$

Since i_p is directly proportional to the light-flux F, equation (3) may be written

$$F = \lambda \{\exp(-\beta V')\}\{\exp(eV'/kT) - 1\} \qquad \text{......(4)},$$

where λ is a new constant. Equation (4) represents an important relation between the light flux and the voltage developed on open-circuit in a rectifier photoelectric cell *in the absence of any applied voltage.* In § 5 will be found examples of its application to the results from a number of specific cells, from which it will be seen that the measured values of the constant β may be very different from the value 1·5 deduced by Frenkel in his one-dimensional treatment. Departures from equation (4) are discussed in § 5.

λ

§ 3. DETAILS OF CONSTRUCTION OF SELENIUM-SULPHUR RECTIFIER PHOTOELECTRIC CELLS

3·1. The cells referred to in § 1 were originally made from steel discs ground flat on one surface, 5·1 cm. in diameter, and 0·8 mm. thick, by coating them with a selenium-sulphur mixture. After the annealing of the layer a thin film of silver was sputtered on to the surface of the mixture. Experiments were continued with these steel discs. Discs of zinc, brass, copper, tin and aluminium also were used, but no really successful cell was obtained; the failure appears to be due to the difficulty of obtaining a small contact gap between the selenium-sulphur film and the surface of these metals. The sulphur was added to the selenium because of the extreme difficulty of coating an iron plate with a thin, uniform, and continuous film of selenium alone; the transformation of the vitreous form of selenium into the grey, metallic variety is attended by a considerable diminution in volume, and this tends to cause rupture at a contact surface and to introduce microscopic cracks and air cavities throughout the material. Thin annealed films of selenium-sulphur are almost entirely free from such mechanical imperfections.

Twenty cells were made with mixtures of selenium and sulphur in various proportions. Pure flowers of sulphur, varying by unit steps from 1 to 10 per cent of the total weight, was added to molten selenium to give ten different mixtures of selenium and sulphur; two cells were made with each mixture. There were, however, no marked variations in sensitivity, but it was found that when the added quantity of sulphur was less than 5 per cent of the total weight the difficulty of obtaining a uniform and continuous film on the iron disc remained. Accordingly subsequent investigation was concentrated upon mixtures containing 10 per cent of sulphur by weight. This mixture melted at approximately 185° C. The melting point of pure selenium is 220° C. The solvent (selenium) separated on cooling, and appeared to be embedded in a matrix consisting of a eutectic mixture of sulphur and selenium. A comparison of the X-ray diffraction patterns given by the pure selenium and by the selenium-sulphur mixture showed that the sulphur atoms entered the selenium lattice; this resulted in a small change in the dimensions of the selenium lattice. The lines given by the selenium-sulphur mixture were sharper than those given by pure selenium; this indicates that the crystals were better developed in the selenium-sulphur mixture than in pure selenium. The selenium used has been described as pure, because consistent success in manufacture was ensured only when the purest selenium available was used. In § 1 Wilson's theory of the rectifier cell was discussed; his treatment depends upon the presence in the semiconductor of impurity sufficient in amount to mask entirely the natural conductivity of the substance. In this work, however, the selenium used was obtained from Germany and was stated to be 100 per cent pure according to chemical analysis; very small traces of metals could be found spectroscopically, but it is doubtful whether these traces of metal represent a sufficient amount of impurity to accord with Wilson's theory.

3·2. *The coating of the iron discs.* Flat circular brass plates, 8 cm. in diameter, were recessed to take one steel disc; plates recessed to depths of 1·2 mm., 1·0 mm. and 0·85 mm. approximately were employed. A brass plate holding one steel disc with its ground surface uppermost was maintained at a temperature of approximately 200° C., while the steel disc was coated uniformly with the molten mixture by means of a flat piece of steel maintained at the same temperature, the surface of the coating being made flush with the surface of the brass plate. With the coated disc in position the brass plate was then cooled rapidly, and while the selenium-sulphur mixture was still plastic its surface was rolled smooth by means of a polished silver-steel roller to give a film of approximately uniform density and thickness. During the rolling, the ends of the roller were in contact with the flat edge of the brass plate. Selenium-sulphur films having approximate thicknesses of 0·4, 0·2 and 0·05 mm. were thus obtained. A slit cut in the edge of the brass plate facilitated the removal of the coated iron disc after rolling.

3·3. *Annealing of selenium-sulphur films.* In the annealing experiments an electric oven with thermostatic control was employed. The temperature was measured by means of a mercury-in-glass thermometer. Experiments on the annealing of the vitreous selenium-sulphur mixture were carried out over a range of temperature of from 140° C. to 185° C. for periods varying from 30 minutes to 30 hours. The results indicate two distinct stages in the annealing. For example, when a coated disc is maintained at 160–165° C. over a period of 3 hours the transformation to the grey, metallic variety takes place; but the surface of the selenium-sulphur still preserves a smooth glazed appearance and has a violet-grey colour by reflected light. The conductivity is very low. If this period of time is extended, or if the annealing is begun at a higher temperature, say 170–175° C., a further change takes place. The surface becomes neutral grey in colour and has the appearance of having a porous structure; its conductivity is relatively very much higher. The annealing required to produce the necessary allotropic change needs careful attention; no success at all is obtained with an unannealed selenium-sulphur film. The annealing determines not only the conductivity of the semiconductor but also the state of the surface. It has been shown that two quite different types of surface can be produced, but there is no sudden change from one type to the other in the annealing. In fact between the two extremes, which have been named type A and type B, many curious results may be obtained. For the pure selenium and pure sulphur used in these experiments, the conditions of annealing can be specified approximately as follows:

Type A: 3 hours at 160–165° C.
Type B: 6 hours at 170–175° C.

The conditions depend upon the purity of the elements used and must not, therefore, be rigidly applied to every new supply of selenium and sulphur. The determination of these types was assisted by resistance-measurements at a definite applied voltage.

3·4. *Sputtered metal films.* The specimens were sputtered one at a time in a bell jar waxed to a brass plate by means of a wax of very low vapour pressure.

The source of supply was a 2000-V. 0·5-A. direct-current generator. The voltage between the electrodes was measured by means of an electrostatic voltmeter, and a milliammeter was connected in series with the supply. All experiments were carried out with a current-density of 10^{-4} A./cm.^{2} at 1400 V. with a vacuum to give a dark space approximately 3 cm. in length. The specimens were placed at the edge of the negative glow, i.e. just outside the Crookes dark space. An aluminium ring with an inside diameter of 4·9 cm. was placed on the selenium-sulphur surface so as to leave an unsputtered rim of selenium about 1 mm. wide.

Sputtering was tried with the following metals: cadmium, lead, tin, zinc, copper, nickel, silver, gold, platinum and brass. The last apparently sputters as brass. Iron and aluminium do not sputter appreciably.

Successful cells were not obtained with cadmium, lead, tin or zinc, principally because of the oxidation of the cathode; thin films of oxide appeared on the surface of the semiconductor, giving interference colours. The formation of selenides of these low-melting-point metals was also suspected. Attempts made with copper, nickel and brass were only partially successful; oxidation of the cathode occurred, but to a much smaller extent. Accordingly, all subsequent experiments were restricted to sputtered films of gold, silver and platinum.

§ 4. DETAILS OF APPARATUS AND METHODS OF MEASUREMENT

4·1. In all the illumination measurements a 1000-W. 100-V. gas-filled projection lamp operating at 50 V. and giving a candlepower of 216·5 candles in a known direction perpendicular to the plane of the filament was used on a 3-metre photometer bench. The colour temperature of the lamp at 50 V. was given as 2360° K.

A holder for the selenium-sulphur photocell, containing eight fine-wire brush contacts arranged equidistantly round the edge of the sputtered metal film and pressing lightly on it, was made; contact with the back of the iron base of the cell was made by a flat-nosed metal screw. By means of the holder each cell could be easily and accurately mounted on a photometer carriage. A vernier potentiometer was used to measure the p.d. across various standard resistances, 1, 10, 100, 1000, 10,000 Ω. and 1 MΩ., in series with the cell. The maximum range of illumination was from 0 to 225 foot-candles. A holder for an adjustable iris diaphragm was also constructed; in those experiments in which the iris diaphragm was employed, the diaphragm was mounted and centred in front of the cell in a plane parallel with the plane of the cell and at a distance of 1 cm. from its surface.

In the determination of the spectro-photoelectrical sensitivity each cell was connected in series with a galvanometer of resistance 15 Ω. and of sensitivity 2000 mm./μA. at 3 metres. A constant-deviation spectrometer was employed with an Adam Hilger linear thermopile made for the spectrometer. No slit-width correction was used. The wave-length band at 5900 A. was approximately 50 A. No precautions were taken to allow for stray light in the spectrometer. Apart from correction for these possible sources of error, the results give the spectro-photoelectrical sensitivity of these cells.

4·2. In the course of this work about three hundred cells were constructed; considerable experience with regard to the general behaviour of these cells under illumination was thus gained. It will be appreciated that with the many possible variables in construction the reproduction of cells with precisely similar properties is a matter of considerable difficulty. This is not surprising in view of the fact that surface effects can still only be treated phenomenologically. For special purposes exactly similar cells may be produced if during the sputtering process the response of the cell is compared with that of a completed cell by illuminating the cell under construction to a known intensity through the bell jar. The sputtering may be discontinued when the two cells show the same response over a small but definite range of illumination.

The precise measurement of the response of a cell presents, of course, no difficulty. Where comparisons are made between the results obtained from cells, among which there are known variations in the controllable variables in construction, the results given apply to cells whose response represents the average response obtained to an accuracy of about ± 1 per cent.

It will be necessary to comment on certain specific cells in § 5; these cells have been numbered for reference purposes, and are specified in table 1.

Table 1

Cell no.	Approximate thickness t of selenium-sulphur film (mm.)	Type of surface	Sputtered metal	Time of sputtering (min.)
1	0·4	B	Gold	23
4	0·2	B	Silver	60
5	$0{\cdot}05 < t < 0{\cdot}1$	A	Platinum	4
6	0·2	A	Platinum	4
7	0·4	B	Platinum	12
9	0·4	B	Platinum	15
10	0·4	B	Gold	30
11	$0{\cdot}05 < t < 0{\cdot}1$	A	Platinum	4

§ 5. APPLICATION OF THEORY AND EXPERIMENTAL RESULTS

5·11. A complete rectifier photoelectric cell may be represented by the equivalent circuit given in figure 2, in which R_f denotes the effective resistance of the sputtered metal film, R_s the bulk resistance of the semiconductor, R_c the effective resistance of the transition layer measured from the metal to the semiconductor, and R_e the resistance of any external measuring instrument. When $R_e \to \infty$, i_r, the reverse current, becomes equal to i_p, the primary photoelectric current; but when R_e is finite, the reverse current is no longer equal to i_p, since some fraction of i_p will be transmitted in its original direction, and a current i_e depending upon the magnitude of R_e will flow in the external circuit.

R_f

R_s, R_c

R_e

i_e

Thus

$$i_e = i_p - i_r' \qquad \text{......(5)},$$

where i_r' denotes the new value of the reverse current. On open circuit $i_e = 0$, and $i_r' = i_p \equiv i_r$.

i_r'

If in any actual cell i_p is independent of R_e, the voltage developed across the transition layer must obviously depend upon R_e, which determines the value of i_r', and upon the value of R_c.

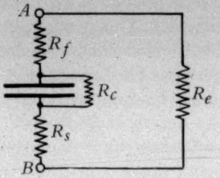


Figure 2. Circuit diagram of rectifier photoelectric cell.

The reverse current i_r' is given by equation (3), and since $i_r' \neq i_r$ when R_e is finite, the voltage developed across the transition layer is no longer equal to its
V_0 open-circuit value V'. Let the voltage developed be V_0 for any given small value of R_e. If R_e is small the variation of the transmission coefficient with variation of V_0 may be neglected; i.e. R_c may be regarded as constant over a limited range of F, if R_e is small. Therefore from equation (1)

$$i_r' = \alpha \{\exp (eV_0/kT) - 1\},$$

or

$$V_0 = \frac{kT}{e} \log \left(\frac{i_r'}{\alpha} + 1\right) \qquad \text{......(6).}$$

But if Ohm's law is assumed for the resistances R_f, R_s and R_e,

$$V_0 = i_e(R_f + R_e + R_s) \qquad \text{......(7).}$$

Hence from equations (6) and (7),

$$\frac{i_e e\,(R_f + R_e + R_s)}{kT} = \log \left(\frac{i_r'}{\alpha} + 1\right) \qquad \text{......(8).}$$

In (8) put $\frac{e}{kT}(R_f + R_e + R_s) \equiv A$. Then

$$i_r' = \alpha \{\exp (Ai_e) - 1\} \qquad \text{......(9).}$$

The value of i_r' obtained from (9) may be substituted in (5), whence

$$\begin{aligned} i_e &= i_p - \alpha \{\exp (Ai_e) - 1\} \\ &= i_p - \alpha \left\{Ai_e + \frac{(Ai_e)^2}{\lfloor 2} + \frac{(Ai_e)^3}{\lfloor 3} + \dots\right\} \end{aligned}$$

or

$$i_e(1 + \alpha A) = i_p - \frac{A^2\alpha i_e^2}{\lfloor 2} - \frac{A^3\alpha i_e^3}{\lfloor 3} - \dots.$$

$$\therefore\ i_e = \frac{1}{1 + \alpha A}\left\{i_p - \frac{A^2\alpha i_e^2}{\lfloor 2} - \frac{A^3\alpha i_e^3}{\lfloor 3} - \dots\right\} \qquad \text{......(10).}$$

From equations (5) and (10) it is seen that since i_p is directly proportional to F, i_e can never be accurately proportional to F. If the deviation from true proportionality is small when $R_e = 0$, and if R_f and R_s are small, and if B denotes a constant in the proportionality of i_p to F, equation (10) may be rewritten to give to a first approximation the relation between i_e and F, namely

$$i_e \cong \frac{1}{1+\alpha A}\left\{BF-\frac{A^2\alpha B^2}{\underline{|2}}.F^2-\frac{A^3\alpha B^3}{\underline{|3}}.F^3-\dots\right\} \qquad \text{......(11).}$$

This equation holds for all the cells, both of type A and type B.

5·12. From equation (7) it is seen that

$$\frac{V_0}{i_e}=R_f+R_e+R_s.$$

Since $V_0 \neq V'$, this quotient cannot be determined experimentally owing to the impossibility of separating R_c, R_f and R_s for the purposes of measurement.

5·13. When $V' < 5$ mV., expansion of the exponential terms in equation (3) shows that with sufficient accuracy

$$i_p=\frac{\alpha e}{kT}V' \qquad \text{......(12),}$$

or

$$V'=\frac{kT}{\alpha e}i_p=R_c i_p \qquad \text{......(13).}$$

When V' is very small $R_c \cong R_c'$, the value of the contact resistance when $V'=0$. Thus when $V' \to 0$, $R_c \to R_c'$ and in the limit

$$R_c'=\frac{kT}{\alpha e} \qquad \text{......(14).}$$

From equation (13), $V'=0$ when $R_c=0$, and therefore in the case of a very good contact, defined by $R_c=0$, no voltage is developed on illumination.

5·14. In view of the difficulty of determining without reference to the experimental work exactly what is given by the quotient of the measured open-circuit voltage and the measured short-circuit current for the same value of F, this quotient, often vaguely termed the "resistance of the cell", will be considered in § 5·4.

Table 2

Sputtered metal	Cell type	Time of sputtering (min.)	Open-circuit voltage developed at 224 foot-candles
Gold	B	15	0·01080
Gold	B	23	0·00500
Silver	A	8	0·08440
Silver	A	9·5	0·06730
Silver	A	10	0·06080
Silver	A	12	0·04730
Platinum	A	4	0·18062
Platinum	A	6	0·11230
Platinum	B	14	0·05520
Platinum	B	15	0·03280

5·2. *Variation of open-circuit voltage with intensity of illumination.*

5·21. Table 2 gives the voltage developed on open circuit at an illumination of

224·0 foot-candles by a number of cells sputtered for various times. For selenium-sulphur surfaces of the class A type, experiment has shown that the open-circuit voltage V' for any given illumination increases with the time of sputtering up to a maximum after from four to five minutes. With further increase of the time of sputtering, V' decreases for the same illumination; if the sputtering be continued beyond sixty minutes, V' decreases rapidly to a very low value with increase of thickness of the deposited metal film, which eventually becomes opaque to light after several hours of sputtering. The initial decrease of the open-circuit voltage from its maximum with increase of the time of sputtering cannot be, however, attributed solely to increased absorption of light by the metal film. Equation (13) shows that V' decreases with decrease of R_c'; it is therefore probable that the time of sputtering determines not only the light-transmission of the film, but also the resistance of the contact formed with the semiconductor.

It is a matter of common experience in the molecular deposition of metals by sputtering that different types of surface receive different amounts of metal at the same voltage, current-density and pressure of vacuum, and in the same time. The B type of surface proves to be more receptive than the A type; but more important is the fact that it has proved impossible experimentally to form a cell with a B type of surface which can for the same illumination give as high an open-circuit voltage as that given by a cell of the A type. Therefore R_c' can never be as high for the B type as for the A type. In the latter case the transition layer may be of pure sulphur, which in the former may be partially or wholly removed by prolonged annealing at a higher temperature, which decreases the effective width of the gap δ between the semiconductor and the metal film.

5·22. The intensity of illumination E, varied by changing the distance of the lamp from the cell along a photometer bench, is directly proportional to the light flux F incident on the cell so long as the illuminated area of its surface remains constant. When the whole surface of the cell is illuminated, equation (4) may be rewritten to give the relation between V' and E, thus

$$E=\gamma\,\{\exp(-\beta V')\}\,\{\exp(eV'/kT)-1\} \qquad \text{......(4')},$$

γ, r where $\gamma=\lambda/\pi r^2$, r being the radius of the cell.

The experiments show that the equation (4′) holds for all cells of the A type when β is positive and of the order of 15.

For type A cell no. 5 at 300° K. the observed values of V' for a range of illumination of from 25 to 200 foot-candles, figure 3, are represented to an accuracy of about 1 per cent by the equation

$$E=3{\cdot}41_2\,\{\exp(-15{\cdot}4V')\}\,\{\exp(eV'/kT)-1\} \qquad \text{......(15)}.$$

The experimental value of β is therefore approximately ten times the value 1·5 theoretically deduced by Frenkel and Joffé. Table 3 shows at various illuminations the experimental values of V' and the values calculated from equation (15) for this cell.

It will now be shown that the value of β for any cell depends upon the ratio of the illuminated area to the area within the circular collecting electrode.

5·23. Cell no. 11, of type A, was mounted with the adjustable iris diaphragm on the photometer bench as described in § 4. The smallest aperture available was first used, the incident illumination being varied by moving the lamp along the bench. With each opening two curves were plotted showing (1) open-circuit voltage against light-flux, (2) short-circuit current against light-flux. The diaphragm was opened to give a range of known areas of aperture, the diameter of each opening being measured.

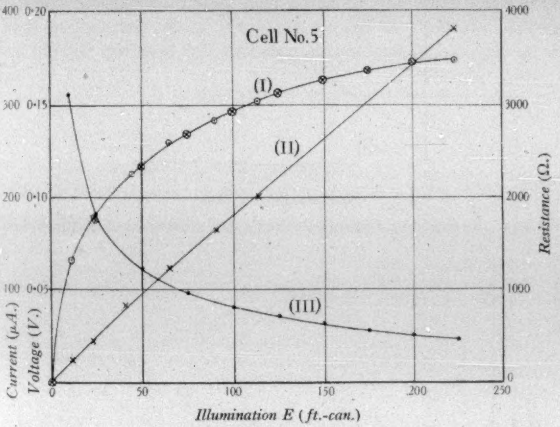


Figure 3. Illumination against (I) voltage, (II) current and (III) resistance.

Table 3

Illumination, E (foot-candles)	Open-circuit voltage, V', experimental values ($T=300°$ K.)	Open-circuit voltage, V', values calculated from equation (15) ($T=300°$ K.)
25	0·0885	0·0875
50	0·1165	0·1160
75	0·1340	0·1330
100	0·1460	0·1460
125	0·1560	0·1560
150	0·1630	0·1630
175	0·1680	0·1680
200	0·1730	0·1735

The value of β for each curve of open-circuit voltage against light-flux was then determined; these values are plotted against radius of aperture in figure 4. The short-circuit current sensitivity, expressed in μA. per lumen, was, for a given flux, found to be constant to within 2 per cent for any value of the aperture-radius; but the open-circuit voltage for the same flux increased with decrease of the radius, figure 5, and it was found that the value of β depended upon the size of the aperture. Extrapolation to zero radius in figure 4 shows that for an extremely small radius

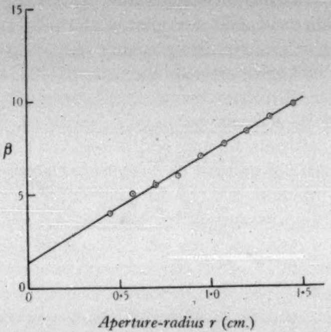


Figure 4. Curve relating β (equation 1) and radius of aperture of iris diaphragm.

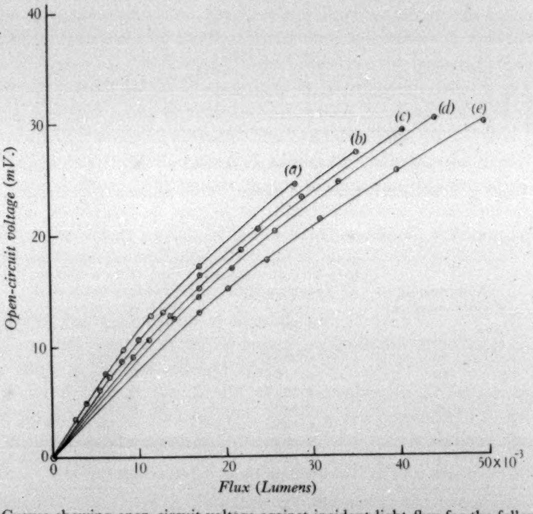


Figure 5. Curves showing open-circuit voltage against incident light-flux for the following values of aperture-radius (cm.): (*a*) 0·575, (*b*) 0·825, (*c*) 1·075, (*d*) 1·325, (*e*) 1·45.

(corresponding to Frenkel and Joffé's one-dimensional treatment) the experimental value of β would be equal to 1·4, which is quite close to Frenkel and Joffé's deduced value 1·5. This interesting result can be readily interpreted in terms of the effect of space-charge limitation of current within the illuminated area of the metal film. The effect is, of course, absent in the one-dimensional treatment of Frenkel and Joffé.

The maximum open-circuit voltage measured for any given flux is limited by the number of photoelectrons collected by the circular electrode round the periphery of the sputtered film. In the absence of mechanical imperfections in the film there will be a radial distribution of electron-flow, but the electrons travel against a retarding potential dependent upon their initial position in the film and due to the space charge of electrons arising between them and the electrode. Electrons with insufficient kinetic energies will therefore return through the gap to the semiconductor.

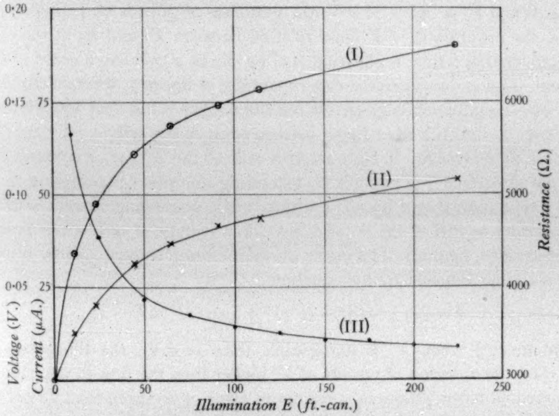


Figure 6. Curves showing illumination against (I) voltage, (II) current and (III) resistance for cell no. 6.

Equation (15) for class-A cell no. 5, with the whole surface illuminated, may be rewritten to give the relation between the measured open-circuit voltage V' and the primary photocurrent that arises from, and is directly proportional to, the illumination E. Thus

$$i_p' = B\,\{\exp(-15\cdot4V')\}\,\{\exp(eV'/kT)-1\},$$

where B is a constant.

But if Frenkel and Joffé's theoretical equation for the one-dimensional case held for this cell, then for any value V' of the measured open-circuit voltage the true primary photocurrent i_p'' collected by the circular electrode *in the absence of space-charge limitation* would be given by

$$i_p'' = \alpha\,\{\exp(-1\cdot5V')\}\,\{\exp(eV'/kT)-1\}.$$

It follows from these two equations that

$$i_p' \propto i_p'' \exp(-13\cdot9V') \qquad \text{......(16)}$$

for any value V' of the measured open-circuit voltage. This relation between i_p' and i_p'' is similar in form to the corresponding relations in the theory of thermionic emission. The practical effect of this space-charge limitation is to make the transmission coefficient for electron-flow from semiconductor to metal appear to decrease more rapidly with V'. In other words, the contact resistance R_c from metal to semiconductor decreases more rapidly with increase of V' than it would in the absence of a space-charge limitation of current, and the voltage shows a saturation effect. But as the illuminated area is reduced symmetrically about the centre of the cell the space-charge effect will be less important and the value of β calculated from the observations should fall, approaching the theoretical limit 1·5; this is in accordance with our observations.

5·24. When V' is below 10 mV. the omission of powers of V' higher than the first (i.e. the assumption of a linear relation between V' and E) in the expansion of the exponential terms in equation (15) produces a calculated error of less than 4 per cent. This is demonstrated experimentally in figure 7, where a curve relating voltage with illumination is given for cell no. 6 (type A) at very low illuminations, a 1 per cent neutral-tint filter being used in front of the cell.

5·25. *Cells of type B.* In contrast with cells of the A type, it is found that with the B type of surface it is possible by extending the time of sputtering to produce cells of a type represented by no. 4 which, while possessing current-sensibility of the same order as that of the A type, develop a voltage of only a few millivolts at 100 foot-candles, figure 8. The curve of open-circuit voltage against illumination for cell no. 4 is expressed by the equation

$$E = 399{\cdot}4\,\{\exp(-10{\cdot}7V')\}\,\{\exp(eV'/kT) - 1\} \qquad \ldots\ldots(17).$$

In equation (17) when V' is not greater than 10 mV., the illumination being 169 ft.-can., the omission of powers of V' higher than the first in the expansion of the exponential terms produces a maximum error of no more than 10 per cent.

It will be noted that γ in general equation (4′) is, for cell no. 4 of type B represented by equation (17), more than 100 times the corresponding value for the class-A type represented by equation (15). The long time of sputtering produces a reduction of the metal-film resistance with increase in the thickness of the film; and if the light-transmission of the sputtered metal film in cell no. 4 were only approximately $\frac{1}{100}$ that of no. 5, γ would be increased accordingly. In that case, however, the selective absorption of the thicker film in cell no. 4 would give rise to a large change in spectro-photoelectric sensitivity; it will be seen later that the spectro-photoelectrical sensitivity curve for cell no. 4 is not very markedly different from that for cell no. 5. The increase in γ is therefore not solely due to increase in thickness of the metal film; from equation (14) it follows therefore that the increase in γ corresponds to a reduction in the value of R_c' for zero V'.

The B type of contact evidently yields a transition layer of a much smaller gap-width δ than the A type.

5·3. *Variation of short-circuit current with intensity of illumination.*

5·31. By measuring the p.d. across a number of standard resistances closing the cell circuit in turn, the short-circuit current, corresponding to zero external resistance, can be found by extrapolation. Equation (11) shows that the relation between the short-circuit current and the intensity of illumination may be written

$$i_e = aE - bE^2 - cE^3 \qquad \ldots\ldots(18),$$

where a, b, c are constants.

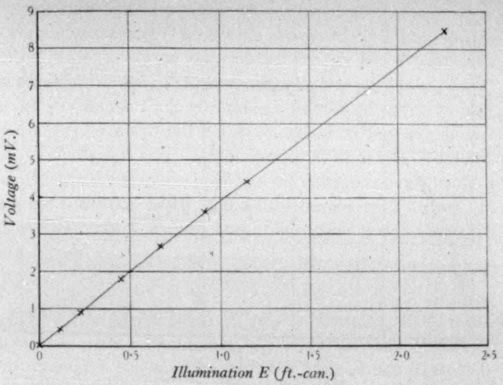


Figure 7. Curve for low illuminations, cell no. 6. Percentage transmission of neutral-tint filter = 1·012.

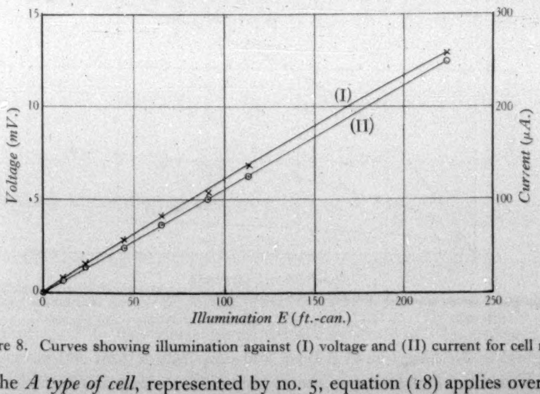


Figure 8. Curves showing illumination against (I) voltage and (II) current for cell no. 4

For the *A type of cell*, represented by no. 5, equation (18) applies over a range of values of E from 10 to 225 foot-candles to an accuracy of 4 per cent; when $a=2$, $b=1\times10^{-3}$, and $c=2\times10^{-6}$, i_e is given in microamperes. It is evident that the range of measurement is too large for the assumption of a constant transmission coefficient, on which the above equation was based, to be precisely true. Over a smaller range of E, such as 10 to 100 ft.-can., equation (18) can be taken as representing accurately enough the relation between i_e and E.

Further, in § 5·11 the assumption was made that Ohm's law holds for the resistances R_f, R_e and R_s. The validity of this assumption is confirmed by comparing the curves of short-circuit current against illumination for cells nos. 5 and 6, figures 3 and 6. Cell no. 6 is of A type, and only differs from cell no. 5 in having a thicker film of semiconductor, so that it gives rise to a larger value of R_s. For cell no. 5 when $R_e = 2772\ \Omega$., figure 10, the current-against-illumination curve is not very different from the short-circuit current-against-illumination curve for cell no. 6, figure 6. Thus in comparison with cell no. 5, no. 6 behaves as if it contained internally (in R_s) an extra fixed resistance of 2772 Ω.

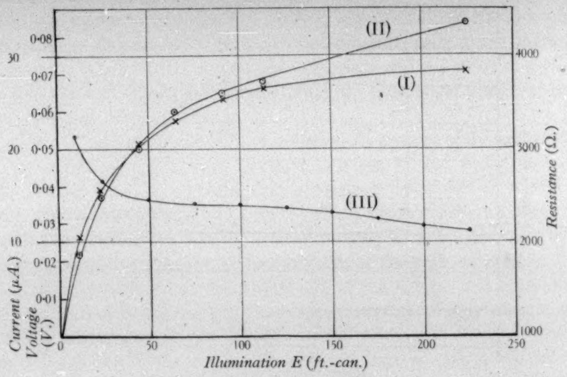


Figure 9. Curves showing (I) voltage, (II) current and (III) resistance against illumination.

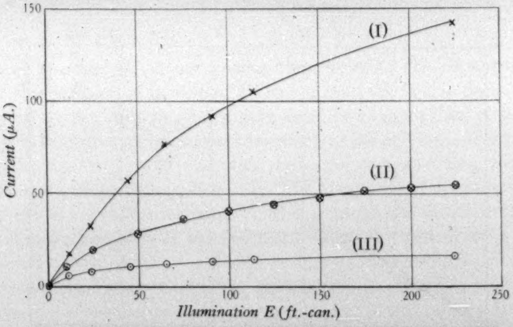


Figure 10. Curves showing illumination against current for various resistances in series with cell. (I) 1000 Ω.; (II) 2772 Ω., calculated current; (III) 10,000 Ω.

For the B type of cell, represented by no. 4, figure 8, the coefficients b and c in equation (18) are so small that the deviation from true proportionality over the whole range 10 to 200 ft.-can. is negligible, i.e.

$$i_e \propto E \qquad \text{......(18')}.$$

5·32. If the constructional details relating to the types of cell represented by nos. 4 and 5 are varied, it is possible to construct cells whose response to illumination is very complicated. These cells can be of little practical use; figure 9, which gives the response curves for cell no. 7 (type B), is included to show the necessity for adhering to the constructional details specified in table 1.

5·4. *The cell resistance.* It is a common practice to term the quotient of the measured open-circuit voltage and the short-circuit current at a given illumination the "resistance" of the cell at that illumination. It must not be assumed, however, that at any given illumination the open-circuit voltage is equal to the short-circuit voltage across the transition layer. Let V' denote the open-circuit voltage, V_0 the short-circuit voltage (when $R_e = 0$), and i_e the short-circuit current at any given illumination E. Then

$$\frac{V'}{i_e} = (R_f + R_s)\frac{V'}{V_0} \qquad \text{......(19)},$$

since $i_e(R_f + R_s) = V_0$. If therefore V'/V_0 is constant, V'/i_e is also constant. Experimentally it is always found that V'/i_e depends on E. From equation (19), therefore, the quotient V'/V_0 must depend on E, since R_f and R_s are to be considered as constant.

Experimentally, it has been found that for the A type of cell, represented by no. 5, the equation

$$\frac{V'}{i_e} = 6150 \exp(-14{\cdot}0V') \qquad \text{......(20)}$$

holds over a range of E from 25 to 200 ft.-can.;

$$\therefore\quad (R_f + R_s)\frac{V'}{V_0} = 6150 \exp(-14{\cdot}0V')$$

and

$$V_0 = \frac{R_f + R_s}{6150} V' \exp(-14{\cdot}0V') \qquad \text{......(21)}.$$

For the B type, represented by cell no. 4,

$$\frac{V'}{i_e} = 58{\cdot}1 \exp(-9{\cdot}0V') \qquad \text{......(22)}$$

and

$$V_0 = \frac{R_f + R_s}{58{\cdot}1} V' \exp(-9{\cdot}0V') \qquad \text{......(23)}.$$

The quotient V'/i_e has been determined as a function of V' in order to give the relation between V' and V_0; V'/i_e is given in ohms and termed R in the figures.

It is seen from equation (18) that as E is decreased, i_e becomes more nearly proportional to E. When E is very small, i_e may with sufficient accuracy be taken equal to i_p. The value of the quotient V'/i_e then approaches R_c', equation (14), the contact resistance at zero voltage. Therefore, from equation (20) for cell no. 5, $R_c' = 6150\ \Omega$.; from equation (22) for cell no. 4, $R_c' = 58{\cdot}1\ \Omega$.

From equation (14) $\alpha \propto 1/R_c'$, and of course γ is directly proportional to α, so that $\gamma \propto 1/R_c'$. From equation (17) for cell no. 4, $\gamma = 399{\cdot}4$; and from equation (15) for cell no. 5, $\gamma = 3{\cdot}412$. The ratio of the γs for the two types is therefore $399{\cdot}4/3{\cdot}412$ or 117·1; on the other hand, the ratio of the reciprocal of the corresponding contact resistances is $6150/58{\cdot}1$ or 105·9. The constant γ is also inversely proportional to the light-transmission of the metal film, and it follows therefore that the light-transmission of the metal film in cell no. 4 is approximately 90·5 per cent that of the metal film in cell no. 5.

The quotient V'/i_e we interpret to represent the apparent variation of the contact resistance with V' arising from the space-charge limitation of the current collected on open circuit by the circular electrode round the periphery of the metal film.

5·5. *Spectro-photoelectrical sensitivity.*

5·51. The spectro-photoelectrical sensitivity curves for unit incident energy for cells nos. 1, 4 and 5 are given in figure 11. The spectro-photoelectrical sensitivity for light of any given wave-length is defined as the current at that wave-length, relative to the current at the wave-length of maximum sensitivity, for the same amount of incident energy.

It is important to note how small are the differences between the three cells. This is at first sight surprising, since cell no. 1 was sputtered with gold for 23 minutes, no. 4 with silver for 60 minutes, and no. 5 with platinum for 4 minutes. A feasible explanation of the smallness of the differences is that the light-transmission of these metal films may be so high that they are not appreciably selective through the wave-length range 4000–8000 A.

5·52. Methods of measuring the thickness of these metal films were investigated with the object of determining whether the tentative explanation given in 5·51 could be supported experimentally. Unfortunately no direct measurements of their spectral absorption can be made. A metal film sputtered on glass would bear no relation, either with regard to the state of the deposit or the thickness of the film, to one deposited in the same time and under the same conditions on a semiconductor surface. Thus it is necessary to seek an indirect method of determining the approximate thickness of these films.

Any experimental method not utilizing the film as deposited on the selenium-sulphur surface is open to objection. The method actually adopted and described below gives a maximum estimate of the average thickness. Two cells, nos. 9 and 10, were selected. J. J. Thomson⁽⁴⁾ gives a formula for the resistance R of a circular lamina of thickness t so small that the currents are compelled to flow parallel to the lamina, being led in and out of the lamina by circular electrodes, radii a and b, placed on its circumference. This formula is

a, b

$$R = \frac{\sigma}{\pi t} \log_e \frac{r^2}{ab} \qquad \text{......(24)},$$

r where r, assumed large compared with a or b, is the distance apart of the electrodes,
σ and σ the specific resistance of the lamina.

It was soon found that the surface conductivity of the selenium-sulphur could be neglected. Accordingly, the optically flat end-faces of two circular steel rods were pressed on to the circumference at the ends of a diameter of the circular metal film. The current flowing when a voltage of 176·5 mV. was applied across the electrodes was measured. The resistances obtained were 400 Ω. for cell no. 9 and 290 Ω. for cell no. 10.

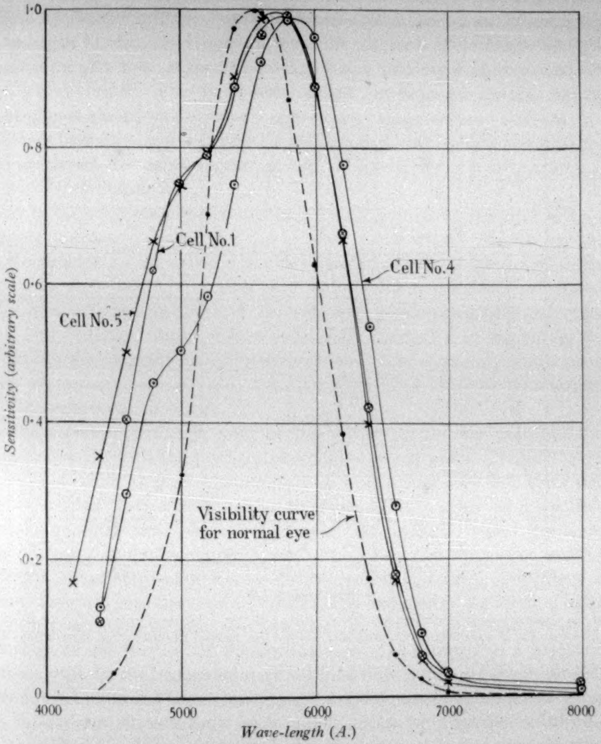


Figure 11. Spectro-photoelectrical sensitivity curves for unit incident energy: gold, silver and platinum films.

Now $\frac{1}{\pi}\log\frac{r^2}{ab}=1{\cdot}99$, so that equation (24) becomes

$$t=\frac{1{\cdot}99}{R}\sigma \qquad \text{......(25).}$$

σ_p
σ_g

For cell no. 9, $t=4\cdot975\times10^{-3}\times\sigma_p$; for cell no. 10, $t=6\cdot86\times10^{-3}\times\sigma_g$, where σ_p and σ_g are the specific resistances in very thin films of platinum and gold respectively.

An examination of published figures$^{(5)}$ for the specific resistances of thin metal films of gold and platinum was made, and those adopted for σ_p and σ_g are the highest possible values, viz. $\sigma_g\cong50\times10^{-6}$, $\sigma_p\cong150\times10^{-6}$.

With these assumed values, the *maximum* average thicknesses of the metal films in cells nos. 9 and 10 are $t\cong7\times10^{-7}$ cm. for cell no. 9, and $t\cong3\times10^{-7}$ cm. for cell no. 10.

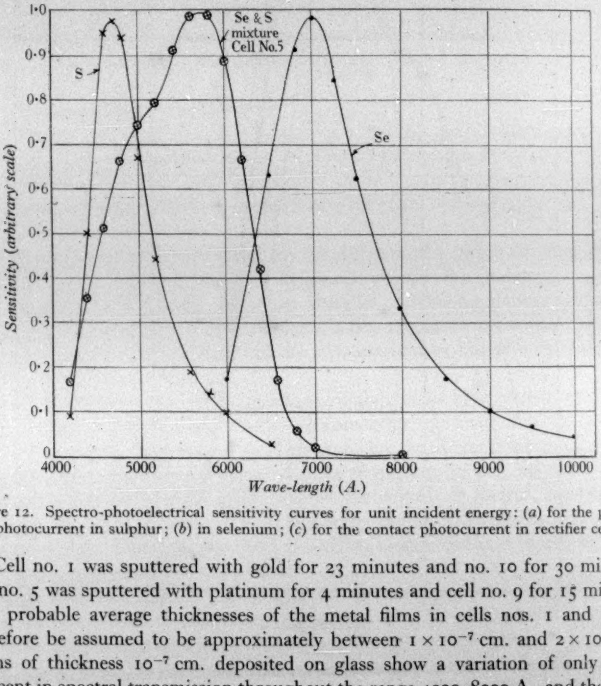


Figure 12. Spectro-photoelectrical sensitivity curves for unit incident energy: (*a*) for the primary photocurrent in sulphur; (*b*) in selenium; (*c*) for the contact photocurrent in rectifier cell.

Cell no. 1 was sputtered with gold for 23 minutes and no. 10 for 30 minutes; cell no. 5 was sputtered with platinum for 4 minutes and cell no. 9 for 15 minutes. The probable average thicknesses of the metal films in cells nos. 1 and 5 may therefore be assumed to be approximately between 1×10^{-7} cm. and 2×10^{-7} cm. Films of thickness 10^{-7} cm. deposited on glass show a variation of only a few per cent in spectral transmission throughout the range 4200–8000 A., and the mean transmission is of the order of 90 per cent according to the International Critical Tables. The thickness of the film in cell no. 4, sputtered with silver for 60 minutes, is greater (see § 5·4) than that of the films in cells no. 5 and no. 1; but it must still be less than 10^{-6} cm. for the explanation given in 5·51 for the smallness of the differences between the spectro-photoelectrical sensitivity curves in figure 11 to hold.

5·53. In figure 12, the spectro-photoelectrical sensitivity for the internal primary photocurrents in selenium and in sulphur (obtained from Gudden's curves$^{(6)}$ and

Kurrelmeyer's curves(7) respectively) is plotted in addition to the spectro-photoelectrical sensitivity for cell no. 5. It is seen from the curves that the threshold energy for the contact photocurrent appears to be greater than for the internal primary photocurrent. Frenkel observed this to be true of the cuprous-oxide rectifier photoelectric cell also, but inferred from his theory that the contact effect actually began at the same frequency as the internal effect, and only became appreciable at a certain higher frequency which decreased as the width of the contact gap increased. The small but marked displacement towards the longer wave-lengths of the curve for cell no. 4 in figure 11 as compared with that for cell no. 5 cannot therefore be due, if Frenkel's theory is correct, to the known differences in the widths of the contact gaps; it can, however, be reasonably ascribed to absorption by the silver film.

5·54. In the selenium photoconducting cell the number of electrons which must be set free in order to account for the increased conductance of selenium under a given illumination far exceeds the number which could be set free by photoelectric action of the absorbed light according to the quantum equivalence principle. The theories of the light-sensitivity of the selenium photoconducting cell remained obscure and doubtful, until Gudden and Pohl(6) succeeded in accounting for the observed changes in conductance by observations of the secondary effects following primary ionization; they proved that secondary photocurrents arise which are essentially electrolytic in nature.

In the rectifier cells dealt with in this paper the current produced can be associated directly with the primary photoelectrons, as in the case of the photoemissive photoelectric cell.(8) Qualitatively all that is necessary is to show that the maximum measured current for any given incident flux is appreciably smaller than the maximum theoretical photoelectric yield. Accordingly if each point on the spectral-energy distribution curve(9) of the tungsten-filament source is weighted according to the spectro-photoelectrical sensitivity of cell no. 5, integration of the resultant curve will give the total effective energy-emission of the lamp. Thus from the known area of the cell the effective incident energy can be calculated for an illumination of 200 ft.-can. If this energy were concentrated in a narrow band of wave-lengths in the region of maximum sensitivity (5700 A.) and completely converted, the electrical yield according to the quantum equivalence principle would be of the order of 6500×10^{-6} A. For cell no. 5, the current obtained at 200 ft.-can. is only 343×10^{-6} A. It is therefore reasonable to conclude that the observed photocurrent is directly associated with the primary photocurrent arising from the optical release of bound electrons. The assumption has already been stated in the discussion of theories in § 2.

If the current density in R_s is very high, one must expect the cell to show all the known characteristics of selenium resistances, namely drifting, inertia and remanence effects, and such have been observed with high values of the incident illumination. For accurate work, therefore, it is important to restrict the use of the cells here described to illuminations giving a short-circuit current not exceeding 100 μA.

§ 6. SEALING OF CELLS

Absorption of moisture is found to lead to disintegration of the sputtered film and to complicated behaviour of the selenium resistance R_s. Work is in progress on the use of moisture-proof colourless lacquers. The results given in this paper were obtained as soon as possible after construction of the cells, which were kept in a desiccator over phosphorus pentoxide.

§ 7. ACKNOWLEDGMENTS

The author is indebted to Mr L. H. McDermott, B.Sc., A.R.C.S., D.I.C. for his assistance in the determination of the spectro-photoelectrical sensitivity, to Mr C. J. W. Grieveson, B.Sc., M.A. for his assistance in the experiments with the iris diaphragm, and to Mr W. A. Wood, M.Sc. and Dr G. Shearer of the Physics Department for the X-ray examination of the selenium-sulphur mixture. He also wishes to record his sincere thanks for the kind interest and encouragement of Dr J. W. T. Walsh and of Mr J. S. Preston, M.A., F.Inst.P. and particularly for much valuable criticism by Dr D. Owen.

REFERENCES

(1) Uljanin, W. von. *Ann. Phys.*, Lpz., **34**, 241–73 (1888).
See also G. P. Barnard, *The Selenium Cell*, p. 32 (1930). (Constable and Co., Ltd.)
(2) Barnard, G. P. *Illum. Engr*, **26**, 163 (1933).
(3) Nordheim, L. *Phys. Z.* **30**, 177 (1929).
Georgeson, W. *Proc. Camb. phil. Soc.* **25**, 175 (1929).
Schottky, W. and Deutschmann, W. *Phys. Z.* **30**, 839 (1929).
Wilson, A. H. *Proc. roy. Soc.* **133**, 458 (1931); **134**, 277–87 (1931–2); **136**, 487–98 (1932).
Condon, E. V. *Rev. mod. Phys.* **3**, 43 (1931).
Frenkel, J. and Joffé, A. *Phys. Z. Sowjet.* **1**, 60–87 (1932).
Frenkel, J. *Wave Mechanics* (1932). (Oxford: Clarendon Press.)
Bornstein, M. *Phys. Z. Sowjet.* **2**, 28–45 (1933).
Fowler, R. H. *Proc. roy. Soc.* **141**, 56–71 (1933).
(4) Thomson, J. J. *Elements of Electricity and Magnetism.* (Cambridge University Press.)
(5) Bartlett, R. S. *Phil. Mag.* **5**, 848–59 (1928), and references.
(6) Gudden, B. and Pohl, R. *Z. Phys.* **35**, 243–59 (1925).
(7) Kurrelmeyer, B. *Phys. Rev.* **30**, 893–910 (1927).
(8) Hughes, A. L. and Du Bridge, L. A. *Photoelectric Phenomena* (1932). (McGraw Hill Book Co.)
(9) Skogland, J. F. *Misc. Publ. U.S. Bur. Stand.* No. 86 (1929).

DISCUSSION

Dr Walsh. The type of cell which the author has studied is of great and continually increasing importance to those who are interested in the measurement of illumination and in photometry generally. It is therefore of great value to have a description of the way in which these cells can be made, so that the mystery which has hitherto surrounded their construction no longer daunts anyone wishing to adapt them to a particular problem or type of measurement.

For illumination photometry the two matters of outstanding importance are (i) the spectral sensitivity curve, and (ii) the extent to which the sensitivity of the cell varies with the obliquity of the incident light. The ideal would be a cell which was unaffected by obliquity and for which the sensitivity curve was exactly the same as that for the eye. It is greatly to be hoped that the author will continue his researches and so bring this ideal nearer to realization than it is at present.

Dr M. BENJAMIN. There are one or two points in the paper on which I am not quite clear. How exactly is the current carried through the selenium-sulphur film? Is it electronic or electrolytic in nature? The theoretical treatment in the paper makes use of Wilson's theory of semi-electronic conductors. This theory depends on the presence of an impurity in the pure substance, the impurity atoms supplying electrons to the empty energy-band of the pure substance, these electrons then becoming conduction electrons. If the theory is applicable to these rectifying cells, it would appear that the current should mainly be conducted electronically.

The author states that spectroscopic evidence indicates that the amount of any impurity is probably insufficient to allow Wilson's theory to apply. Actually the theory will apply even when the impurity-concentration is only of the order of one part in one million. Does the author believe that the conductivity is at all dependent on the presence of a suitable impurity, and if so, has he any views as to its nature? With regard to the marked dependence of the conductivity on the annealing temperature of the sulphur-selenium film, may it not be that the exact nature of the sulphur-selenium lattice is dependent on the annealing temperature? Alterations in the lattice will result in alterations in the energy levels, and therefore in the conductivity.

The X-ray examinations showed that mixed crystals of selenium and sulphur were formed. Did the chemical composition agree exactly with the amount of sulphur added, or is the composition dependent on the annealing temperature? If the solid solution does vary in composition, this may explain the variations in conductivity.

AUTHOR's reply: I should like to thank Dr Walsh for his comments on my work. Dr Benjamin's first question concerning the nature of the electrical conduction in the selenium-sulphur film is answered in paragraph 5·54. Owing to the extreme difficulty of determining both the nature and the amount of the impurities present in the elements used for these cells, any discussion as to how the conduction electrons are derived can have little scientific value. Undoubtedly, the formation of the selenium-sulphur lattice is greatly influenced by the annealing temperature. This has been studied in the case of selenium alone, and the results are considered in my book *The Selenium Cell*. The formation of the eutectic mixture of the selenium and sulphur depends, of course, upon the annealing temperature.

771·351

THE DEVELOPMENT OF THE PHOTOGRAPHIC LENS

By W. TAYLOR and H. W. LEE, Research Department,
Taylor, Taylor and Hobson, Ltd., Leicester

Address delivered on January 18, 1935.

§ 1. INTRODUCTION

This address has been prepared with two main objects. The first is to give British physicists a better conceit of their national heritage, based on the knowledge that the principal fundamental inventions concerned in the development of photographic lenses have been British inventions. The second is to show that, with one exception, the development of photographic lens design has consisted at each successive step of only a small advance on traditional design.

This process of development, which is usual with inventions in general, is well illustrated by the case of the shot gun in its descent from the bow and cross-bow. When gun-powder was invented the cross-bow was in use, and the barrel of the shot gun was fastened to the old stock of the cross-bow; but because it was not seen how to use the trigger, this remained for a time out of use, like the buttons on the back of our tail coats. A touch-hole was put in the side of the barrel so that the charge could be exploded by a match applied to the mouth of the hole. Then, the better to ensure ignition, the mouth was enlarged into a pan which had to be separately primed, and presently a hinged cover was fitted to keep the priming dry. Gradually the trigger came into use again; a spring-urged cock was fitted to carry a flint, and the trigger held this ready for action. When the trigger was released, the flint came down, lifted the cover from the pan, and struck a spark which ignited the priming.

In the course of time the percussion cap was invented, the flint was discarded and the cock became a hammer to strike the separate percussion cap. Then the charge of gunpowder was put in a packet for better convenience and to save time in loading the gun, but the packet had to be broken to expose the powder. This was the beginning of the cartridge. Presently it was seen that it would be better to attach the percussion cap to the cartridge, and in order to reach the cap each cartridge was provided with a pin projecting from its side ready to be struck by the hammer in the gun; but the gun had to be altered to receive this cartridge and became a breech loader. The cartridges were dangerous things to carry, and had to be placed in the gun in a particular way with the pin toward the hammer, and in time the pin was put in the axis of the cartridge. Afterwards it was seen that it was superfluous and dangerous to provide each cartridge with a pin, so the pin was housed once for all in the gun and the percussion cap set flush in the head of the cartridge as we now know it; but for a long time the hammer of the gun operated transversely to the axis of the gun, and finally it was changed to operate along the axis in line with the pin.

The development of the photographic lens has followed a similar course and with few exceptions has clung closely to tradition.

§2. FORM OF SURFACE

The ordinary photographic lens has hitherto employed spherical surfaces, mainly for the reason that no other form of surface has been commercially attainable with the requisite accuracy. The order of accuracy commonly desirable and commonly attained is within one or two wave-lengths of light, measured as departure from sphericity by means of an accurate counterpart of the surface placed in contact with it so as to exhibit any disconformity by means of Newton's rings. Lens-surfaces of such accuracy are, and probably must always be, produced by grinding and polishing with abrasive charged laps; and the only form of tool which can be moved upon the work in various directions to abrade it while maintaining continuous surface contact with it so as to govern its shape is the spherical form.

No appliances based upon mere line or point contact have hitherto been made, nor probably ever will be made, sufficiently rigid to operate commercially with this accuracy.

§3. HISTORICAL

The first step. The use of simple lenses with spherical surfaces for forming images of objects goes back to antiquity. It was Chester Moore-Hall an English amateur who in 1733 discovered, and Dollond an English optician who in 1757 independently re-discovered and "reduced to practice", the fact that by suitably combining a double-convex collective lens of crown glass with a nearly plano-concave dispersive lens of flint glass, it was possible to correct the chromatic error of single lenses which causes different colours of the spectrum to focus at different distances. They also discovered that, by the same means, it was possible to correct the so-called spherical aberration of the single lens which causes rays of light from one object point, incident at various zones of the lens surface, to focus at a corresponding variety of distances instead of at one common distance. The angular extent of the image well defined by Dollond's objective was only about one or two degrees, but this invention furnished the common form of telescope objective used for over a century, and still used; and this became the nucleus from which more and more perfect lens systems for photography were evolved.

The second step in this evolution was made in 1830 by Lister, the father of Lord Lister, who constructed microscope objectives by combining two of Dollond's telescope doublets (the two flints facing the object which he wished to magnify) to obtain increased aperture and illumination, and he so spaced them apart that they corrected each other's coma.

The third step was collateral with Dollond's invention. In 1802 Wollaston, an English physician and amateur optician, discovered how to correct astigmatism and produce lenses giving reasonably well-defined images of greater angular extent. He applied his invention first to the correction of the astigmatism of spectacle lenses, and in 1912 to the camera obscura and camera lucida. He perceived that in order to be free from astigmatism the lens must focus at the same distance lines radial in the field and lines at right angles thereto; and this he secured, as shown in figure 1, by giving the lens a meniscus form and placing a stop at a suitable distance from its concave side. The significance of the distance was that it determined the particular zone of the meniscus through which light passed at any given angle to the lens axis. Probably Wollaston arrived at his discovery experimentally, but it readily follows from consideration of the first-order aberrations.

E, I, d If E is the coefficient of spherical aberration, I that of coma, and d the distance of the stop from the lens, the condition for removal of astigmatism in a simple lens with a stop is

$$Ed^2+2Id+1=0 \qquad \ldots\ldots(1).$$

For a collective lens, E is positive and d^2 necessarily so; so that Id must be negative, i.e. the coma will be of different sign according to whether the stop is placed in front of or behind the lens. In the case of spectacles the stop or diaphragm is the centre of rotation of the eye, and is consequently on that side of the lens remote from the object. A century elapsed before the advantages of Wollaston's meniscus spectacle lenses were generally appreciated, but the form is now in common use.

In the camera obscura Wollaston placed the lens between the stop and the focal plane or, as we should say, placed the lens behind the stop, so that d is positive in equation (1); consequently I must be negative. It can easily be shown that equation (1) always yields two real solutions if the lens is a meniscus with its concave surface facing the stop. Of course, in practice the smaller value of the stop-distance d is used.

If the lens is placed in front of the diaphragm, d is negative and I must be positive. It can be shown, by putting in the value of the aberrations in terms of the lens form, that there will be real roots only if the ratio of the power of the first surface to that of the lens is greater than the square of the refractive index. This relation seems to have been overlooked for a century, and from it a construction of meniscus lens has recently been patented(1).

The locus of all image points in the field of Wollaston's lens (which he called periscopic) was not a plane but a surface concave to the lens—in current terms, his lens had an uncorrected field; also it was not corrected for spherical or chromatic aberration or distortion. At small apertures, however, the definition was sufficiently good for early photographic cameras and this simple meniscus of Wollaston is still made by the million for cheap cameras.

Of *the fourth step*, the correction of distortion, the authorship is disputed, but h it is based on Wollaston's invention and dates from about 1860. Distortion h can be expressed in terms of the other aberrations by an equation which, for a simple lens, has the form:

$$h=(3+1/n)\,d+3I.d^2+E.d^3 \qquad \ldots\ldots(2)$$

and this cannot be zero unless $d=0$. It follows from the equation that if two lenses are equally spaced on opposite sides of a stop, and have equal powers F and equal amounts of coma but of opposite signs, the distortion will be small if E, the coefficient of spherical aberration, is not large, since E is multiplied by the cube of a small quantity d. These conditions are nearly satisfied by employing two equal menisci on opposite sides of a diaphragm. Such a system is also approximately free from coma. It is still used in photography, generally under the name of "the periscopic system". The aberrations still uncorrected are spherical and chromatic aberrations and field curvature, and these limit the aperture to about $f/11$.

The next problem to arise was that of increasing the aperture of the lens in order to make portrait photography commercially feasible. This, *the fifth step*, was first taken by Joseph Petzval, a Hungarian mathematician, in about 1840. His lens in shown in figure 2. Starting with the objective of Dollond, corrected for spherical and *chromatic* aberrations, he placed a stop behind to correct its astigmatism as Wollaston had done. He chose a form with rather a small value of coma, but this entails a large stop-distance, as equation (1) shows. Behind the stop he placed another pair of lenses and at such a distance that their astigmatism was corrected by the stop. These pairs of lenses were *separately corrected* for spherical and chromatic aberrations and the rear pair had coma equal but of opposite sign to that of the front pair (whence the difference in form), so that the comas cancelled out. His complete system was thus corrected for spherical and chromatic aberrations, coma and astigmatism, and he realized an aperture of $f/3$ (i.e. the diameter was one-third of the focal length).

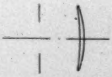


Figure 1. Wollaston's meniscus lens, 1812.

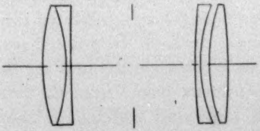


Figure 2. Voigtländer (Petzval) portrait lens, 1840.

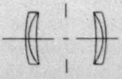


Figure 3. Dallmeyer rapid rectilinear, British patent specification no. 2502/1866. A. Steinheil aplanat, Bavarian Pat. 1866.

Petzval's lens had the disadvantage that curvature of field was uncorrected and further that, owing to the great distance of its members from the stop and from one another, its field was very small (with a total angle of about 15°). Nevertheless, in the hands of the portrait photographer (a great conservative) this lens largely survives.

In the 'sixties Dallmeyer in England and Steinheil in Germany, concurrently, took *the sixth step* by introducing the rapid rectilinear lens which gave the photographer a lens with a moderately extended field. The rapid rectilinear (or rapid symmetrical or rapid aplanat, as it was variously called) is shown in figure 3. It also started from the inventions of Dollond and Wollaston, and employed two cemented *achromatic* doublet objectives, each with a large amount of coma and consequently requiring only a small stop-distance to correct the astigmatism, equation (1). The two faced one another so that one was the mirror of the other with respect to the stop, and thereby coma and distortion were corrected, see equation (2)—hence the name "rectilinear". The large amount of coma was obtained through suitable choice of glass. If cemented achromatic doublets free from spherical aberrations are calculated for a number of different pairs of glass (crown and flint), it will be found that those made from pairs having small differences in refractive index and in dispersion have larger amounts of coma than those made from pairs having greater difference in these optical properties.*

The inventors of the rectilinear lens employed light flint glass for the crown lens, together with the usual dense flint. The smallness of the difference in refraction and dispersion between these two glasses produced a doublet having a large amount of coma, and of meniscus shape with very deep curves. The small stop-distance determined by the coma made the system short, and the lens had the comparatively large field of 40° to 50°, but the depth of the curves limited the aperture to $f/8$, and it could not have freedom from astigmatism and a flat field at the same time.

In 1827 G. B. Airy, Astronomer-Royal, indicated the requirements of the next, *the seventh step*, the flattening of the field. He pointed out that the locus of the focus of points off the axis was different when the rays are in the plane containing the object point and lens-axis (primary plane) from the locus when the rays are in a plane perpendicular thereto (secondary plane). The two loci he found for a single lens (considering up to the second power of the angle of field) to be circles both concave to the lens and having the respective radii of curvature $nf/(3n+1)$ and $nf/(n+1)$, where f is the focal length and n the refractive index. If astigmatism is corrected the two loci coincide with a radius of field curvature of nf. Airy's work was unknown on the continent and thirteen years after his publication (in 1840) the condition for flatness of field was independently discovered by J. Petzval.

Instead of considering the radii of curvature as Airy did, it is usual to consider the curvature of the field, viz. $1/nf$ for F/n, where F is the power of the lens. Then for a system of thin lenses the total curvature of the field is the sum of the values

F_1, F_2 F/n. In particular for a pair of glasses, as in an achromatic doublet, if F_1 and F_2 are

v_1, v_2 the respective powers and the glass materials are defined as n_1 and v_1 and n_2 and v_2 respectively, the condition for flat field is:

$$F_1/n_1 - F_2/n_2 = 0,$$

and the condition for achromatism is:

$$F_1/V_1 - F_2/V_2 = 0,$$

V, n_1, n_2 V is defined as $(n-1)/(n_1-n_2)$, where n_1 and n_2 are the refractive indices for the two

n rays to be achromatized, and n their mean value.

If these two equations are to be consistent, a large refractive index must be associated with a large value of V or a small dispersion, and glass of low index should have large dispersion, whereas in the crown and flint glasses alone available in the time of Airy and Petzval high index was associated with high dispersion.

In 1888, Abbe and Schott produced glasses with the necessary properties, and soon afterwards Schroeder (of Ross) patented(2) the first anastigmat, that is, a lens having a substantially flat field substantially free from astigmatism. The lens, which is shown in figure 4, consisted of two *achromatic* doublets each using the new barium crown glass of high refractive index for the outer glasses, and the new light flint glass of low refractive index for the inner glasses to correct the field; astigmatism and coma were corrected as in Wollaston's invention by suitably spacing the stop in relation to the lens members. Correction of spherical aberration was impossible with this form of lens, as it would require the refractive index of the flint to be higher than that of the crown. The aperture was accordingly limited to *f*/16.

---

* This has been discussed for example in *Constructional Data of Small Telescope Objectives*, by T. Smith and R. W. Cheshire.

Thus it seemed that the new glasses had removed one aberration, field-curvature, at the expense of another, spherical aberration. Rudolph (of Zeiss) then conceived the idea, shown in figure 5, of combining an *achromatic* doublet in front made of old glasses, which would correct the spherical aberration, with one in the rear made of the new glasses, to flatten the field. This in 1890(3) was the first anastigmat corrected for spherical aberration. The old achromat was in front and the new achromat behind the diaphragm. The aperture was *f*/7·5, and slightly better than that of the rectilinear lens. Four years later(4) he combined his old and new achromats in one, making a quadruplet which he duplicated. This is shown in figure 6.

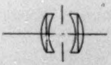


Figure 4. Ross (Schroeder) concentric, British patent specification no. 5194/88.

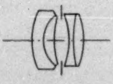


Figure 5. Zeiss (Rudolph) protar, British patent specification no. 6028/90.

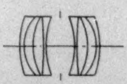


Figure 6. Zeiss (Rudolph) double protar, British patent specification no. 19,509/94.

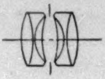


Figure 7. Goerz (v. Hoegh) dagor, British patent specification no. 23,378/92.

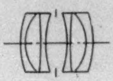


Figure 8. Zeiss (Rudolph), British patent specification no. 4692/93.

The order of refractive indices in each member was: low-index flint, high-index crown (forming the new achromat), low-index crown, high-index flint (forming the old achromat). Two of these quadruplets were used, one before and one after the diaphragm, to correct coma and distortion. The new achromats were the inside pairs and the old the outside pairs of the complete system of eight glasses.

Von Hoegh (of Goerz) in 1892(5) accomplished the same result with only six elements, figure 7. He placed the pairs of new glasses outside instead of inside the pairs of old glasses so that the order of refractive indices was such that in each half of the system he could give two interior elements the same index, and these became one single double-concave dispersive element, which was set between two collective elements. Zeiss afterwards adopted the same simplification with a similar sequence of refractive index, figure 8. His components consisted each of a double-convex element between a double-concave element next the stop, and a meniscus dispersive element outside. But this form was abandoned by Zeiss in favour of the earlier eight-element type.

§4. THE BREAK FROM TRADITION

We have seen that up to this time the practical development of the photographic lens had proceeded step-by-step along lines which had become traditional and which tended to more and more elaborate constructions. In the earlier stages Englishmen had taken a predominant part, not only in invention but in the mathematical treatment of the subject. Airy's work was extended in 1829 and 1830 by the Cambridge mathematician Coddington, whose work was in turn the basis of Dennis Taylor's and was developed by him in his *Applied Optics* published in 1906; but latterly the Germans had taken the lead, and this was due partly to the theoretical work of Von Seidel who had, in 1856, reduced the mathematical expressions for optical aberrations to forms more convenient for the use of lens-computers than any previously published in the German language, and partly to the stimulus given by the glass production of Abbe and Schott.

In 1893 another Englishman, with the penetrating vision of genius and the true spirit of inventiveness, stepped aside from tradition and solved this complex problem by using only three lens elements. This was H. Dennis Taylor who was honoured for his work by receiving the Duddell Medal of this Society last year. He satisfied the Airy-Petzval condition in a new way; see figure 9. He began by considering a collective and a dispersive lens of equal or approximately equal power, and spaced apart so that they had a combined collective power to form a real image. If two lenses of powers F_1 and F_2 are separated, and the weaker (if they are not equal) is dispersive, then the system is collective and the greater the separation the greater the power; but the Airy-Petzval sum, indicating the degree of curvature of the field, remains the same whatever the separation. Consequently, by employing a negative and a positive lens separated from one another, it is possible to correct the curvature of field with almost unlimited choice of refractive indices.

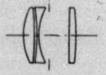


Figure 9. Taylor-Hobson (H. D. Taylor) Cooke lens (final form), British patent specification no. 15,107/95.

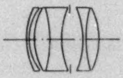


Figure 10. Dallmeyer (Aldis) stigmatic, British patent specification no. 16,640/95.

In order to correct distortion he divided the postulated collective component into two, placing one part on one side of the dispersive component and the other on the other side. Hitherto it had been considered necessary to achromatize each separate component of a lens system in order to achromatize for points off the axis. Dennis Taylor, although in his first design he followed this tradition, subsequently discovered that he could achromatize the system as a whole without achromatizing the individual components. He corrected spherical aberration and coma by suitable choice of curves, and astigmatism by making the separation from the stop of the collective components greater than that of the dispersive component, i.e. he placed the collective elements outside the dispersive element, and the stop near to the last.

No fundamentally new principle of photographic lens design has been originated since Dennis Taylor invented this lens forty years ago.

H. L. Aldis (then with Dallmeyer), in his "stigmatic" lens, figure 10[(6)], changed the order of Dennis Taylor's elements by placing the dispersive outside the collective components. This necessitated, in all components, a meniscus form for the correction of spherical aberration and the compounding of some or all of them for chromatic correction.

§5. INFLUENCE OF THE CINEMA

During the latter half of this period the predominant influence in the development of photographic lenses has been that of the cinema and film production; and this has demanded lenses substantially free from all aberrations and of increasingly large apertures.

At first when lenses with apertures so large as $f/2$ were offered to the photographers in Hollywood they were rejected because experience had taught that apertures must be limited in order to get sufficient depth of focus; but this objection has been removed by the realization that for a given size of final image the shorter the focal length of the lens used for the original photograph the greater is its effective depth of field, and owing to the small area of the cine film picture (the "frame") relatively short-focus lenses with larger apertures have come into use. Thus the glare and heat of excessive illumination are reduced, greatly to the comfort of the actors.

The recent development of fine-grain photographic emulsion for cinema film has stimulated a general demand for smaller cameras and lenses of proportionately short focal length, greater consequent depth of field, greater rapidity, and generally more critical definition.

In these and other developments, since the date of Dennis Taylor's invention, three principal types or families of lens have been the subjects of work by various opticians. First the continental type of lens based on the designs of Rudolph (of Zeiss), figure 6, and Von Hoegh (of Goerz), figure 7, which are capable of yielding relatively large field but with restricted aperture. Second, the Petzval type, figure 2, characterized by extreme aperture but small field; and third, lenses based on the principles of Dennis Taylor's invention and characterized as being capable of yielding both extreme aperture and moderately large field.

§6. THE CONTINENTAL TYPE OF LENS

Lenses of the continental type had one advantage in the eyes of photographers, namely their convertibility. Since they were made of two similar components of different focal length, separately achromatized, either component could be used by itself as a partially corrected lens and with these the photographer had two, or sometimes three, different lenses at his disposal. Continental design was directed either to improving the performance of the separate half-lenses or to increasing the aperture, both aims generally leading to greater complexity of design. Messrs Watson increased the aperture of the Zeiss six-lens $f/6{\cdot}8$ convertible lens to $f/4{\cdot}8$, while Messrs Ross (Mr Hasselkus) increased the aperture $f/6{\cdot}8$ of the Zeiss eight-lens formula to $f/5{\cdot}5$ in their "combinable" lens[(7)], while Taylor, Taylor and

Hobson (Mr Lee)(8) further extended the aperture to *f*/4·5. Ross (Hasselkus and Richmond) increased the aperture, *f*/6·8, of the Goerz lens to *f*/4 by separating one element in each component, figure 11(9). This lens, on account of its large field (70°) and aperture, has proved valuable for photographic aerial survey work. Other modifications were made by Steinheil and Voigtländer, figure 12.

The greater complexity resulting from efforts directed to increasing the convertibility of this type of lens is shown in lenses made by Goerz, figure 13(10), Turner and Reich(11) and Taylor-Hobson (H. W. Lee), figure 14(12).

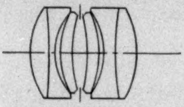


Figure 11. Ross (Hasselkus and Richmond) wide angle, *f*/4, British patent specification no. 295,519.

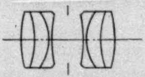


Figure 12. R. Steinheil orthostigmat, British patent specification no. 12,949/95. R. von Voigtländer collinear, British patent specification no. 18,723/96.

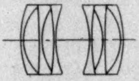


Figure 13. Goerz (v. Hoegh), British patent specification no. 13,904/97.

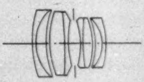


Figure 14. Taylor-Hobson (H. W. Lee) convertible, British patent specification no. 376,044.

§ 7. THE PETZVAL TYPE OF LENS

To supply the demand for more rapid lenses the Petzval type of lens has been developed, and Mr Warmisham (of Taylor, Taylor and Hobson) has invented many modifications and improvements of this lens, taking the aperture to *f*/1·5. The form illustrated in figure 15 has proved particularly suitable for use in the photographic recording of sound, on account of its field correction. Figure 16 shows a modified Petzval lens by Kodak. Zeiss, by adding another element, figure 17, have recently extended the aperture to *f*/0·85 to make possible the cinematography of X-ray skiagrams, work specially associated with the names of Drs Russell Reynolds in England and Robert Janker of Bonn. These high-speed lenses retain the characteristic qualities of the original Petzval lens, namely narrow angle of field and field curvature, faults which become aggravated as the aperture is increased.

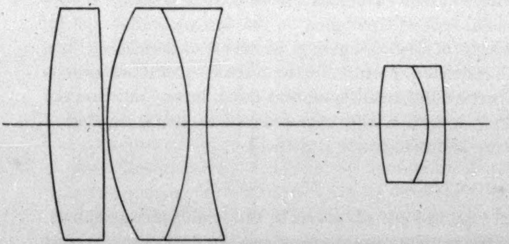


Figure 15. Taylor-Hobson (A. Warmisham), British patent specification no. 342,889.

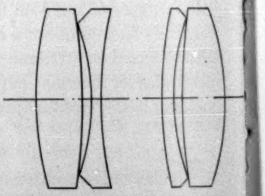


Figure 16. Kodak (Frederick and Altmann), U.S. patent specification no. 1,620,337.

§ 8. DESCENDANTS OF THE H. D. TAYLOR TYPE OF LENS

With the continental type of lens restricted in its aperture, and the Petzval type in its field, the invention of Dennis Taylor has become more and more dominant and the subject of development by various opticians.

Dennis Taylor's lens of three simple elements can, by the use of glasses made available since the date of his invention, be made capable of performance over a field of 40° at an aperture of *f*/3 or, in the case of short-focus lenses, an aperture of *f*/2·5.

He had shown that while only two lenses, collective and dispersive, were necessary for field-flattening, it was necessary to divide the collective lens into two for the sake of other aberrations. He also suggested that the dispersive lens could be divided into two, and these elements could be placed one on each side of the stop and between the two collective lenses as shown in figure 28. These two forms of Dennis Taylor's invention, which may be called the "triplet" and the "quadruplet", have been the parents of later developments which have generally consisted in (*a*) replacing one or more of the simple elements by more or less complicated cemented components; (*b*) replacing one or more single elements by a pair or pairs of elements of like kind (for example as the quadruplet arose from the triplet), and (*c*) combinations of (*a*) and (*b*).

Many of these modifications have resulted in improved performance of the type, and some of them have resulted in avoiding patent claims! Before illustrating examples of these elaborations of the original Dennis Taylor lens, we illustrate, in figure 18, a modification due to H. L. Aldis in the cementing together of the front pair of elements. We have selected for illustration a number of characteristic developments of the Dennis Taylor triplet.

In the class (*a*) may be mentioned the Zeiss tessar (Rudolph and Wandersleb), figure 19, in which the back component is double though not achromatized, and the Voigtländer heliar, figure 20, in which both front and back are made double. L. B. Booth also compounded the front and back to produce the Dallmeyer *f*/2·9 lens, figure 21.

Recent developments, figure 22, of the triplet are due to Berek (of Leitz)(13) who, by making all the components complex, realized an aperture of *f*/2; and, figure 23(14), apertures *f*/2·8 to *f*/2, to Zeiss.

In the second class (*b*) comes the Taylor-Hobson *f*/2·5 lens, figure 24, with double-back component, and the Ernemann (now Zeiss-ikon) series of lenses having apertures up to *f*/2, figure 25, in which the front lens was first divided into two separate simple elements but is now compounded, though not achromatized, and falls into class (*c*), figure 26. Figure 27 shows an example of divided dispersive component, the Meyer plasmat *f*/1·5. The most useful form of this class (*b*) is that in which both dispersives are double-concave, in accordance with the original suggestion of Dennis Taylor. This quadruplet was developed by Goerz in their celor and dogmar lenses, and further, figure 28, by Warmisham (Taylor, Taylor and Hobson) for photography from aircraft during the war(15), for which purpose he improved its corrections for oblique spherical aberration and coma and for zonal astigmatism, so

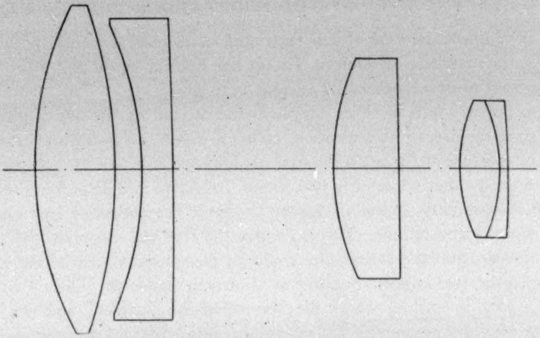


Figure 17. Zeiss *F*/O 85, U.S. patent specification no. 1,967,836.

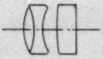


Figure 18. H. L. Aldis, British patent specification no. 5170/01.

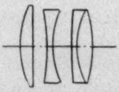


Figure 19. Zeiss (Rudolph and Wandersleb) tessar, British patent specification no. 13,061/02.

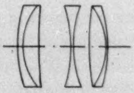


Figure 20. Voigtländer (Harting) hilia, British patent specification no. 13,441/02.

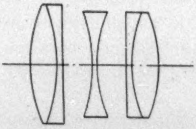


Figure 21. Dallmeyer (L. B. Booth) pentac, British patent specification no. 151,506.

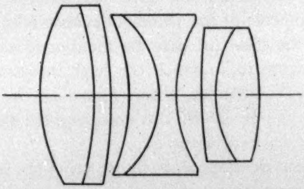


Figure 22. Leitz (Berek), British patent specification no. 382,246.

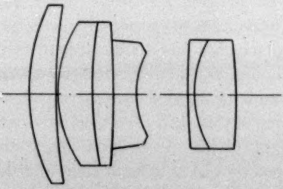


Figure 23. Zeiss, British patent specification no. 383,591.

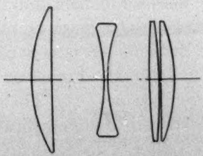


Figure 24. Taylor-Hobson (H. W. Lee), British patent specification no. 224,425.

that a very small aberration patch for points off the axis resulted. This type of lens has been developed by Dr Ross in the United States of America and by H. W. Lee, in another direction, for the photography of stars for measuring their positions. In a lens made for the Cape Observatory of 35 inches focal length and including an angle of 40°, the mounting is interesting because it provides for free circulation of air so as to reduce the effects of differences of temperature on the system. The requirements of such lenses are that the star-images should be perfect circles, that is to say that the images of all points in the field must be quite free from coma and astigmatism, for the position of a star-image can be accurately measured only if the image is circular.

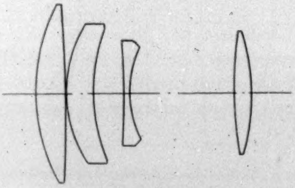


Figure 25. Ernemann, British patent specification no. 237,212.

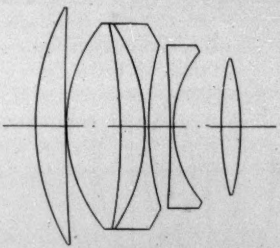


Figure 26. Ernemann, British patent specification no. 237,519.

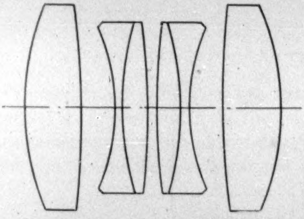


Figure 27. Meyer (Rudolph) plasmat, British patent specification no. 401,630.

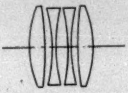


Figure 28. Taylor-Hobson (A. Warmisham) aviar, British patent specification, no. 113,590.

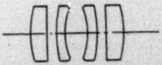


Figure 29. Busch omnar (K. Martin), British patent specification no. 19,504/01.

The quadruplet can be put into another form by making all the separate components simple meniscus elements. This was done in 1900 by H. Meyer(16) and in 1901 by K. Martin (of Emil Busch)(17) with an aperture *f*/6·8, figure 29. Lenses of this general form have since been extensively made by other manufacturers. When modified, it is particularly adapted to provide wide-angle lenses limited to small apertures.

Rudolph (of Zeiss) employing thicker dispersive components, figure 30, was able while satisfying the Airy-Petzval condition, to use shallower curves, and thereby the aperture was increased to *f*/3·5. This was the planar(18) which comes into class (*c*). In 1920 H. W. Lee(19) increased the aperture to *f*/2, figure 31. Recently

Zeiss(20) have increased the aperture to *f*/1·4, and Taylor, Taylor and Hobson(21) (H. W. Lee) to *f*/1·5, while more recent constructions of Taylor, Taylor and Hobson, shown now for the first time, are due respectively to A. Warmisham, figure 32 (aperture *f*/1·3)(22) and H. W. Lee(23), figure 33 (aperture *f*/1).

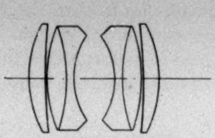


Figure 30. Zeiss (Rudolph) planar, British patent specification no. 27,635/96.

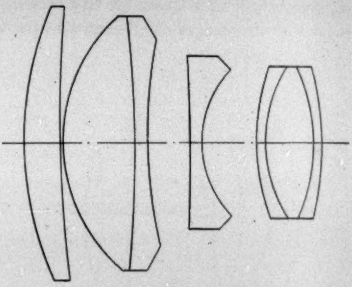


Figure 32. Taylor-Hobson (A. Warmisham), British patent specification no. 408,787.

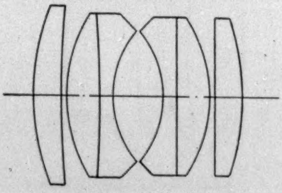


Figure 31. Taylor-Hobson (H. W. Lee), British patent specification no. 157,040.

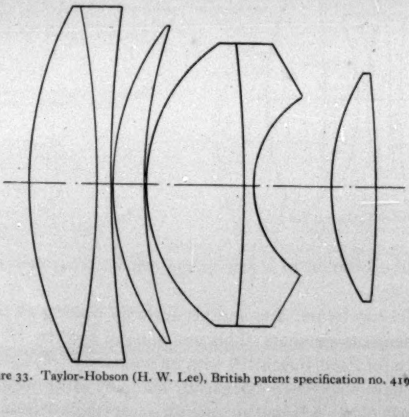


Figure 33. Taylor-Hobson (H. W. Lee), British patent specification no. 419,552.

§9. LENS OF VARIABLE FOCUS

Another problem arising from the demands of the cinema is the provision of a lens of continuously variable focal length. The cinematographer likes to vary the scale of his picture. This can be done discontinuously by applying to the camera a turret of lenses of different focal lengths, or continuously by placing the camera on wheels and moving it about. But the advent of the sound film has rendered movement of the camera during action objectionable. The best solution of the problem is the provision of lenses of continuously variable focal length. In addition to others, references to which are lacking, Warmisham(24) has designed the lens illustrated in figure 34. The difficulties of design are both optical and mechanical. The optical system comprises three relatively movable parts of which the outside two, *A* and *B*, are dispersive and compound, and the inside part *C* is a large-aperture photographic lens. The lens system is drawn in continuous lines to indicate the position corresponding to its long focal length, and the members *A* and *B* are drawn also in dotted lines to show the positions which give the short focal length. When the lens is giving a short focal length there is a large angular field to be corrected, while in the long-focus position the aperture will be small unless the diameters of the lenses are made large, and such diameters would be enormous for the short-focus wide-angle position. Some compromise of aperture with long focus is therefore unavoidable. Mr Warmisham's lens has an aperture of *f*/3·5 at a focal length of 40 mm. and *f*/8 at a focal length of 120 mm.; a 3:1 ratio of scale is thus possible.

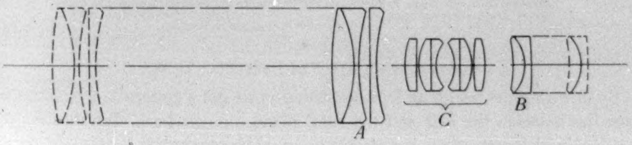


Figure 34. Taylor-Hobson (A. Warmisham) varo, British patent specification no. 398,307.

The mechanical difficulties are concerned with the cutting of precision cams to control the movements of the several components of the lens so as to maintain both the correction of aberrations and the focus on the film, as the focal length changes, and also to alter the iris diaphragm so that the *f* values remain correct.

§ 10. THE TELEPHOTO LENS

The telephotographic lens consists of a collective lens with a dispersive lens at such a distance behind it that the nodal planes are thrown right in front of the system, and the space between the lens and the focal plane is much reduced. Hence while the focal length may be large the camera is compact. The principle was first applied to telescopes in 1834 by Barlow and to photography in 1891, almost simultaneously in England (by Dallmeyer), France and Germany. In distinction from the normal type of lens of the time, with its field concave to the lens, the telephoto field was at first convex, its Airy-Petzval sum being over-corrected owing to the great separation between collective and dispersive components. The performance of these lenses was consequently poor. It was not until 1914 that the first anastigmat telephoto lens, figure 35 (aperture *f*/6), was made by Taylor, Taylor and Hobson under the British patent of L. B. Booth(25) and lenses of this construction were largely used in the War by the Royal Air Force.

Lee and Booth improved this lens by reducing the number of air-glass surfaces and increasing the aperture to f/4·5, the later Lee(26), by separating the collective component into two, figure 36, increased the aperture to f/3·3; and by separating the rear elements he corrected the distortion, a result which has not been achieved in any other telephoto lens. Figure 37 shows the Zeiss telephoto tessar lens(27), and figure 38 shows the Ross (Hasselkus and Richmond) telephoto lens(28).

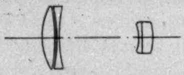


Figure 35. Taylor-Hobson (L. B. Booth) telephoto, British patent specification no. 3096/14.

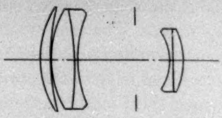


Figure 36. Taylor-Hobson (H. W. Lee), British patent specification no. 222,709.

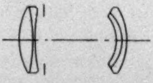


Figure 37. Zeiss tele-tessar, British patent specification no. 179,529.

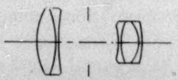


Figure 38. Ross (Hasselkus and Richmond) teleros, British patent specification no. 188,621.

§ 11. INVERTED TELEPHOTO LENS

By inverting the telephoto lens construction we get a system in which the nodal plane lies between the lens and the focal plane, i.e. the back clearance is greater than the focal length. This is sometimes an advantage when lenses of very short focal length are needed; it is a necessity when space has to be provided behind such a lens, for example for the prisms used in some systems of cinematography in colour. This calls for a lens having not only abnormal back focus but also large

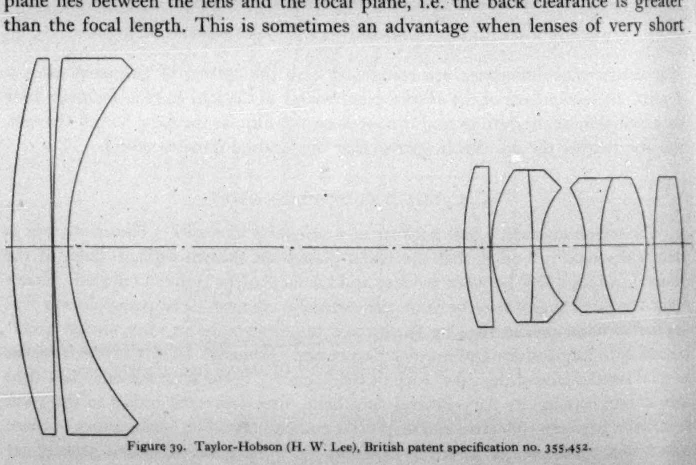
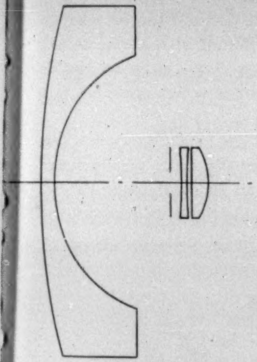


Figure 39. Taylor-Hobson (H. W. Lee), British patent specification no. 355,452.

aperture and considerable field covered at the usual high standard demanded by cinematography. Various forms of inverted telephoto lens have been devised for both projection and photography: for projection by N. Bowen (translux)(29), and W. B. Rayton (of Bausch and Lomb)(30), the aperture being *f*/2·8; and for photography, by H. W. Lee (Taylor, Taylor and Hobson)(31), in a lens having an aperture up to *f*/2, figure 39. The inverted telephoto type of lens has also been adopted by R. Hill and made by Beck(32), figure 40, in a lens for photographing the whole sky for meteorological purposes. The field is 180°, covered of course at an extremely small aperture.



igure 40. R. and J. Beck (R. Hill), British patent specification no. 225,398.

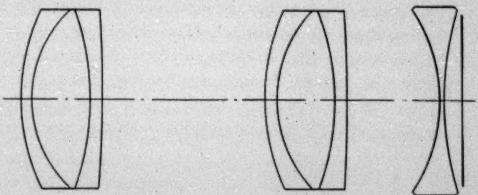


Figure 41. Taylor-Hobson (H. D. Taylor), British patent specification no. 127,058.

§ 12. THE NEGATIVE FIELD-CORRECTOR

In 1866 Prof. C. Piazzi Smyth, the astronomer, suggested placing a dispersive (i.e. negative) lens *close to the focal plane* to correct the curvature of field of the Petzval portrait lens with which he was taking astronomical photographs. In this position the negative lens exerts its full corrective power on the Airy-Petzval sum while it has practically no effect on the focal length or on spherical aberration. So impressed was Piazzi Smyth with the possibility that he constructed a miniature camera, taking pictures 1 in. square, to photograph hieroglyphics inside the Egyptian tombs. This was the first high-speed miniature camera ever made, and was thus some sixty years before its time and suffered the usual fate of being too far ahead of the age. Piazzi Smyth's idea of field-correction was revived by Dennis Taylor in an *f*/2 lens(33), figure 41, made by T. Cooke and Sons and used at Mount Wilson Observatory for stellar photography. The Piazzi Smyth corrector is also now being applied by Dr Ross to the great new astronomical reflecting telescopes to adapt them for photography over a wider field than is otherwise attainable.

§ 13. FUTURE DEVELOPMENTS

This story has now reached the stage when to pursue it entails vision of the future.

No lens is perfect. When we say that aberrations are corrected we mean not that they are reduced to zero but that they are confined within some postulated circle of confusion, and that itself is imperfectly defined.

It seems unlikely that great advance on the present position can be made by adding further complexity to lenses of the types we have described as having become traditional. It may be that another genius (and we may hope again an English genius) may see a new and simplified solution of a problem which again has been pursued near to the limits set by complexity.

It may be that a solution will be discovered in the use of surfaces other than spherical, but that would not become practicable generally until we discovered means of producing and testing them during production and correlating the two surfaces of each element and the various elements with the requisite accuracy.

REFERENCES

(1) British patent specification no. 377,036.
(2) British patent specification no. 5194/1888.
(3) British patent specification no. 56,109/1890.
(4) British patent specification no. 19,509/1894.
(5) British patent specification no. 23,378/1892.
(6) British patent specification no. 16,640/1895.
(7) British patent specification no. 29,636/14.
(8) British patent specification no. 165,657.
(9) British patent specification no. 295,519.
(10) British patent specification no. 13,904/97.
(11) British patent specification no. 9528/1895.
(12) British patent specification no. 376,044.
(13) British patent specification no. 382,246.
(14) British patent specification nos. 350,323 and 383,591.
(15) British patent specification nos. 113,590 and 312,536.
(16) British patent specification no. 20,349/00.
(17) British patent specification no. 19,504/01.
(18) British patent specification no. 27,635/1896.
(19) British patent specification no. 157,040.
(20) British patent specification no. 297,823.
(21) British patent specification no. 397,281.
(22) British patent specification no. 408,787.
(23) British patent specification no. 419,552.
(24) British patent specification no. 398,307.
(25) British patent specification nos. 3096/14 and 139,719.
(26) British patent specification no. 222,709.
(27) British patent specification no. 179,529.
(28) British patent specification no. 188,621.
(29) British patent specification no. 348,123.
(30) U.S. patent specification no. 1,934,561.
(31) British patent specification no. 355,452.
(32) British patent specification no. 225,398.
(33) British patent specification no. 127,058.

DEMONSTRATION

532 . 137

Apparatus for Measuring the Viscosity of Liquids at High Pressures. *Demonstration given on December* 7, 1934 *by* C. C. Mason, O.B.E., M.A., F.Inst.P., *by the courtesy of the* Burmah Oil Company, Ltd.

The formula used in calculating the viscosity from the velocity of the sinker is that due to Lawaczeck. It contains terms showing that the velocity is proportional to the cube of the clearance between sinker and tube, and directly proportional to the difference in density between sinker and liquid. It is thus possible with several sinkers to cover a large range of viscosities. Four are provided having radial clearances of 0·6, 0·8, 1·0 and 1·25 mm. The sinkers are of brass and are hollow, but are made to receive a maximum of four tungsten weights. It is therefore possible to change the effective density, with the result that with the four sinkers viscosities between 10 and 500 centipoises can be determined.

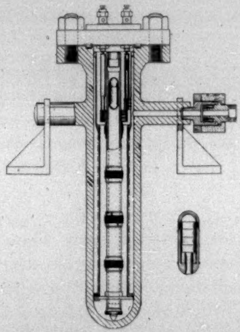


Figure 1. Sectional view of the apparatus.

The inner brass tube is surrounded by another tube inside the steel tube. This intermediate tube carries a resistance mat, and is designed to encourage convectional circulation when current passes through the mat. Connexions to this and to the coils are brought in through the cover by means of an arrangement similar to a sparking plug. A vapour-pressure thermometer shows the temperature and, being provided with an adjustable contact, permits control of the temperature through the medium of a relay.

In the demonstration the image of the galvanometer string was projected on to a screen, and when the instrument was inverted the four deflections were observed as the sinker passed through the four coils.

The instrument is primarily for the purpose of determining the pressure coefficient of viscosity of liquids, and particularly of the mud mixtures used in the boring of oil wells. Heating-arrangements are included in the design in order that temperature effects also may be investigated.

The fundamental principle is not new and consists in measuring the velocity of fall of a cylindrical sinker with spherical ends, under gravity, in a tube. This inner tube is of brass and is quite incapable of supporting the pressure of 4000 lb./in.^{2} which will be used. It is therefore enclosed in an outer steel tube. This is filled with oil and has a connexion to a pressure pump, and the pressures are communicated to the contents of the inner tube through a rubber diaphragm at one end. The outer tube has trunnions so that it can be inverted for the fall of the sinker. The movements of the sinker cannot be followed visually, and this difficulty is surmounted by the method that is used to measure the velocity of a projectile from a gun, although in the present case the velocities and dimensions are small. A small piece of cobalt magnet is inserted in the sinker, and on the outside of the tube are four coils equally spaced 7·5 cm. apart. These are connected in a series-parallel arrangement to a small permanent-magnet string galvanometer, and as the sinker passes through each coil the galvanometer string is deflected. The deflections are recorded photographically by means of a falling-plate camera, time lines being put on the record by momentarily interrupting the light through the galvanometer every tenth of a second with spokes on the rotor of a small synchronous motor driven from a.-c. mains.

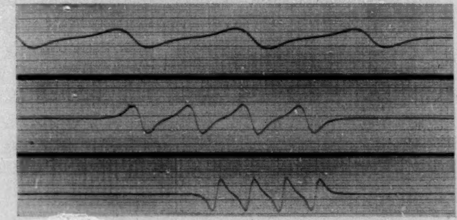


Figure 2. Typical records (reduced to half size).

REVIEWS OF BOOKS

Introduction to Atomic Spectra, by H. E. White, Ph.D. Pp. xii + 457. (New York and London: McGraw-Hill Publishing Co., Ltd.) 30*s*. net.

Practically every account of atomic spectra which has appeared in the last ten or twelve years is either too brief or too highly mathematical and theoretical for those about to take up the subject for special study. Dr White, a member of the team of very distinguished spectroscopists in the University of California at Berkeley, therefore set out to produce for such students an account to which neither of these objections could apply, and he has succeeded most admirably. This excellent book is highly recommended both to beginners and also to those who are already familiar with parts of the subject; and it is the only book of which the reviewer is able to say this so unreservedly and so cordially.

Each section of the subject is opened as nearly as possible at the very beginning, logically developed, and unsparingly illustrated with excellent diagrams and spectrograms. One of the most satisfying features of the work is that the spectra themselves are never lost sight of for longer than is necessary. Short tables of selected observational data are given in the text wherever examples are required for further clarity of the discussion. Any attempt to include exhaustive numerical data, even for one atom or one group of atoms, would obviously be out of place in this new volume of the *International Series*, which already includes Bacher and Goudsmit's recent book of data on *Atomic Energy States*. These two volumes might well be regarded as two parts of a single treatise, almost indispensable to the present-day spectroscopist.

In the well-founded belief that the student can best understand the newer quantum mechanics through a thorough acquaintance with the spectra themselves and the Bohr-Sommerfeld orbits and the vector model, Dr White devotes the opening chapter to the early empirical analysis of line spectra, then develops the older and newer quantum theories of atoms having one valency electron (chapters II–X), and afterwards deals with the more complex spectra of atoms having two or more valency electrons (chapters XI–XV). In the later chapters (XVI–XXI) accounts are given of X-ray spectra, iso-electronic emitters and their spectra, hyperfine structure, perturbations, the Stark effect, and the breadths of lines. The book is well indexed.

The printing too is excellent; in only one respect, so far as the reviewer has yet noticed, could a slightly better setting be suggested: an exponent might be printed after, rather than directly above a subscript; e.g., "$n_2{}^2$" and "$Z_0{}^2$" rather than "n_2^2" and "Z_0^2". Perhaps two other matters in which some readers may not quite agree with the author might be mentioned. First, such an expression as "11 wave numbers" (p. 36) is not a good alternative to "11 wave-number units" or "11 cm.$^{-1}$"—in the analogous case nobody would say "11 wave-lengths" instead of "11 A.U.". Secondly, as to the discovery and naming of the *fundamental series* (pp. 6 and 9) the late Prof. Hicks pointed out in his *Treatise*, and several times in conversation, that *fundamental* (which he first used in the sense *hydrogen-like* or *Balmer-like*) is just a name for which there is neither more nor less to be said than for the other names *principal*, *sharp* and *diffuse*; to call them *Bergmann series* seemed to do scant justice to Saunders's earlier recognition of the Cs doublets or to Runge's and Ritz's discoveries of the connection with the D term. Two or three minor slips also might be mentioned. The last line of p. 25 should read "2 for ionized helium". Of the three sections of Rowland's map mentioned on p. 2, only two appear in figure 1.2; the section including the sodium D lines is omitted; and the section which purports to show the *E* lines (of iron and calcium) is cut off just too short to include these, but has as its most prominent feature the *b* triplet of magnesium, which, moreover, is incorrectly lettered as *E* in figure

1.1; again "*H* lines" in the top line of p. 3 should be "*H* and *K* lines". A second edition will, no doubt, afford Dr White an early opportunity of setting these matters right, for his book will surely be a standard work for many years to come.

W. J.

The Structure of Spectral Terms, by W. M. HICKS, Sc.D., F.R.S. Pp. xi+209. (London: Methuen and Co., Ltd.) 10*s.* 6*d.* net.

The late Prof. Hicks's studies in spectroscopy were directed to the discovery of numerical relations between the wave-numbers of atomic spectrum lines and terms with little or no reference to matters concerning current theory of atomic structure. These studies extended over the last twenty-five or more years of his long and active life. When, in the early summer of last year, he was seized with illness which proved fatal (he died on August 17, within five weeks of his 84th birthday), the proof sheets of this his last work had only just been received and the proof-reading had hardly begun. In this formidable task he was to be assisted by Mr J. R. Clarke, one of his old students and now Lecturer in Physics at the University of Sheffield, from which Dr Hicks retired in 1917. Prof. S. R. Milner therefore joined Mr Clarke in the remaining stages of the work.

In this book Dr Hicks has considerably extended the results set out in his earlier and larger *Treatise on the Analysis of Spectra*, which appeared about the same time as Prof. Fowler's *Report on Series in Line Spectra* and Paschen and Götze's *Seriengesetze der Linienspektren* in 1922 and is now out of print. The opening chapter deals with the representation of line series by formulae, and includes, as a very useful and instructive example, the evaluation of the four constants of the more general of Dr Hicks's two modifications of the Rydberg formula; these two formulae, it may be noted, are still widely regarded as the most useful and accurate of the many which have been proposed for this purpose. In the following chapters a large amount of new observational evidence is presented concerning special features of Dr Hicks's empirical analyses such as satelloids, high-order emission lines, on relations between multiple terms, collaterals, linkages, and summation lines. In the ninth and last chapter atomic structure is discussed in relation to the results of the foregoing analyses. As is to be expected from the very nature of the work and the large amount of numerical material handled, the book is not an easy one to read. It is excellently printed and, in conjunction with his former *Treatise*, forms a fitting record of Dr Hicks's long, painstaking and enthusiastic research in this branch of spectroscopy.

W. J.

Resonance Radiation and Excited Atoms, by A. C. G. MITCHELL and M. W. ZEMANSKY. Pp. viii+338. (Cambridge University Press, The Cambridge Series of Physical Chemistry.) 18*s.* net.

This book, by two workers in the field, fills an obvious gap in the literature of experimental optics. It begins with an introductory chapter on line spectra and the elementary processes involved in resonance radiation, followed by a very useful account of typical light-sources appropriate to their experimental investigation. The second chapter deals with stepwise radiation and with physical and chemical effects of collisions with excited atoms. In chapter III absorption lines are discussed in detail, and an account is given of the measurement of the durations of excited states. Chapter IV, on "Collision processes involving excited atoms", deals rather fully with the broadening of absorption lines and with quenching. The final chapter, on the polarization of resonance radiation, gives a comprehensive account of the effects of magnetic fields, and a brief account of those of electric fields, on the radiation.

The book will be an indispensable work of reference to all who are seriously interested in the subject, and the bibliographies are remarkably up to date. The theory is in places a little too severely compressed, and the theoretical sections are rather dispersed among the experimental descriptions, so that the book will make difficult reading for the unaided beginner, but taken as a whole it is one upon which the authors are to be warmly congratulated and thanked.

H. R. R.

The Diffraction of Light, X-rays and Material Particles, by C. F. Meyer. Pp. 473. (University of Chicago Press.) 22*s.* 6*d.*

This is a book of a rather unusual type; it consists of a detailed exposition of the phenomena of diffraction treated with as little mathematical apparatus as possible. It may be said at once that it is extremely well done; the explanations are clear, the style easy and sympathetic. In addition, the author has given historical information which is much fuller than that usually given in text-books and which adds considerably to the interest. He deals fully with a number of points usually omitted, such as the apparent source of light diffracted from an edge, and his treatment of, for example, the distinction between the Fresnel and Fraunhofer classes of effects, and of Babinet's principle, is unusually thorough and careful. The methods of mounting the plane and concave gratings and the errors of gratings are dealt with at considerable length.

It is probable that an English student who is advanced enough to want such a detailed account of one aspect only of wave theory would be able to grasp a more mathematical treatment and would save time by doing so. On the other hand, many capable students find difficulty in grasping the rather difficult physical principles underlying the mathematics, and for these the book will be very valuable. In addition it should be said that many of the more mathematical questions are dealt with in appendices.

The second part of the book, dealing with X-rays and de Broglie waves, is considerably more detailed, at least on the experimental side, than would be required by most honours students. As the preface explains, it is intended largely for workers in other sciences whose field requires knowledge of these matters, and to this purpose it is well suited. In addition to describing the main experiments on electron-diffraction, the author gives a clear account of the very interesting work of Stern and others on the diffraction of atoms and molecules.

G. P. T.

Newton and the Origin of Colours, by Michael Roberts and E. R. Thomas. Pp. viii + 133. (London: G. Bell and Sons.) 3*s.* 6*d.* net.

It is a pleasure to be able to give an unqualified recommendation to this well-written little book. It makes no pretence of giving an exhaustive account of Newton and his work on colour, but aims rather at a brief sketch of some of the formative influences surrounding his early life, and of the simpler methods of his optical researches. We are given some glimpses of his contemporaries and their comments on his work; and the whole account is used to illustrate the various principles of scientific method and their evolution.

It can be heartily commended as a corrective to any over-dose of technical study, and the present reviewer intends to bring it to the notice of his own students. It is just long enough and sufficiently detailed to create interest, and not so omniscient as to crush discussion. (How often authors fail to hold such a balance!) There are one or two technical queries, as for example "he added to the water *a little* sugar of lead to raise its mean refractive index sufficiently to make it equal to that of glass" (the italics are the present writer's), and the possible misinterpretation of the term "spherical aberration" on p. 38. If indeed it is the case that the aberration now known as "distortion" was originally termed "spherical aberration", some reference in support of this would be of interest. It would be idle, however, to suggest that such insignificant defects as these detract in any way from the success of the book.

If there is any critical feeling left after its perusal, it may be that this "closer" view of Newton's work hardly gives a sense of the real magnitude of his achievements. In a book bearing the above title, some fuller mention might well have been made of his work on colour-mixture, and of the remarkable colour diagram which has formed the foundation of modern methods of colour measurement, and played no small part in the development of the most fruitful theory of colour-vision.

It is doubtful too whether his brilliant achievement in grinding a relatively large-aperture paraboloid has ever received the praise it merits. The figure must indeed have been excellent to allow of the results actually obtained in his reflecting telescope, and little seems to be known of his methods of working mirrors.

However, where so much of interest has been collected in so small a compass, it would be ungenerous to ask for the full discussion of the work of this remarkable man, many of whose sayings bear a freshness and relevance to problems of our own day which is nothing short of amazing.

L. C. M.

Luminescence des Corps Solides, by MAURICE CURIE. Pp. 146. (Recueil des Conférences-Rapports de Documentation sur la Physique, vol. XXIV.) (Paris: Les Presses Universitaires de France.) 50fr.

This recent addition to a deservedly famous series of monographs on various branches of modern physics deals with a vast and fascinating subject to the development of which both physicists and chemists have contributed papers almost innumerable. Comprehensive accounts of it, with extensive bibliographies, have appeared from time to time, e.g. in Kayser's *Handbuch der Spektroscopie*, volume **4** (1908), the National Research Council's Bulletin *Luminescence* (1923), Pringsheim's *Fluorescenz und Phosphorescenz* (1928) and Lenard, Schmidt and Tomaschek's volumes **1** and **2** of the Wien-Harms *Handbuch der Experimentalphysik* (1928). The last-named contains a particularly complete account of the extensive and important work of Lenard and his school. The present report by a notable contributor to knowledge of these phenomena will, without doubt, be very welcome to those who require a concise account of the present state of affairs. It is confined, it is true, to one branch of the subject, namely the luminescence of inorganic solid solutions. These consist, in general, of a crystalline substance, the *diluent*, of low electrical conductivity (such as the sulphides of Zn, Ca, Sr, Ba), containing, as an impurity, a certain, often very small, proportion of the *centre*, i.e. a metal of relatively high atomic number (such as Bi, Cu, Mn, Ni, Pb). Besides this class, of which CaS(Bi) is an example, we also have sulphide mixtures such as ZnS.CdS(Cu), oxides such as CaO(Sm), halides such as KCl(TlCl) and many other compounds. Fluorescence and phosphorescence excited by visible or ultra-violet radiation have been investigated far more extensively than those excited by cathode rays, X-rays, α, β and γ rays, and other means. Accordingly photoluminescence (chapter 1) occupies about two-thirds of the book, and is followed by much shorter accounts of cathodic luminescence (chapter 2), radioluminescence and radio-photoluminescence (chapter 3), triboluminescence and scintillations (chapter 4) and luminescence of solids in flames (chapter 5). A further volume is promised which will treat of the luminescence of gases and vapours, organic substances, and certain inorganic crystalline substances which, unlike those dealt with in the present volume, are in a state of purity.

W. J.

Electrons (+ and −), Protons, Photons, Neutrons, and Cosmic Rays, by R. A. MILLIKAN. Pp. x+492. (London: Cambridge University Press.) 15*s*. net.

Most undergraduate students of physical science have diligently thumbed Prof. Millikan's admirable survey entitled *The Electron*. This volume appeared in 1917, and in revised form in 1924. The present work may be regarded as a third edition of *The Electron*, and, without departing from the historical mode of presentation which gave unity to the older volume the author has been able to make his present volume a thoroughly up-to-date survey of the subject by adding six chapters which deal respectively with waves and particles, the discovery of the cosmic rays, the spinning electron, the positron, the neutron and transmutations of the elements, and the nature of the cosmic rays. The new volume, which gives a view of the position of a most rapidly changing science as it appears at the end of 1934, will be as valuable to the new generation of students as its predecessors were to an older generation.

The name of the author attests sufficiently the manner of presentation.

Actualités Scientifiques et Industrielles. (Hermann and Co., 6 Rue de la Sorbonne, 6, Paris.)

Under the above general heading we have received the monographs listed below. Each is written by an authority on his subject and the treatment is, in general, concise and clear.

115. H. Mineur. *Histoire de l'Astronomie Stellaire.* 15 fr.
116. H. Mineur. *Éléments de Statistique Mathématique.* 12 fr.
141. H. Mineur. *Photographie Stellaire.* 18 fr.
147. Pierre Humbert. *Le Calcul Symbolique.* 10 fr.
159. L. Brillouin. *Les Champs "Self-Consistents" de Hartree et de Fock.* 10 fr.
160. L. Brillouin. *L'Atome de Thomas-Fermi.* 12 fr.
162. P. Swings. *Travaux récents sur les molécules dans le Soleil, les Planètes et les Étoiles.* 14 fr.
177. J. Genard. *Fluorescence des Vapeurs dans le Champ Magnétique.* 12 fr.
181. Louis de Broglie. *Une Nouvelle Conception de la Lumière.* 12 fr.
182. Irène Curie et F. Joliot. *L'Électron Positif.* 10 fr.
185. M. Haïssinsky. *Les Radiocolloïdes.* 9 fr.

Chemical Kinetics and Chain Reactions, by N. SEMENOFF. (International Series of Monographs on Physics, edited by R. H. Fowler and P. Kapitza.) Pp. xii+480. (Oxford: The Clarendon Press, Mr Humphrey Milford.) 1935. 35*s.* net.

The foundations of chemical kinetics, or the theory of the velocities of chemical changes, were laid by the researches of Harcourt and Esson about 1850–66, but the subject was first systematically formulated by van't Hoff in his *Études de Dynamique Chimique* in 1884, who obtained equations for uni-, bi- and multi-molecular changes and gave a simple kinetic interpretation to them. The effect of temperature was expressed in the form $k=Be^{E/RT}$, where k is the velocity constant, B and E are constants, R is the gas constant and T the absolute temperature. Arrhenius in 1889 showed that E may be interpreted as an energy of activation, such that only molecules with an energy exceeding E are capable of undergoing change. Succeeding investigation has, until recently, been mainly concerned with the application and elaboration of these fundamental laws.

As more and more systems were investigated, it became clear that the number of reactions which proceed normally is surprisingly small, and this number is constantly decreasing so that, whereas bimolecular reactions were formerly regarded as the predominant type, a recent monograph on the subject could record only two unequivocal cases of bimolecular reactions.

The idea of a reaction chain appeared about 1913. In one formulation it assumes that the chemical energy set free in a single molecular process ('elementary act') is not directly transformed into heat, but may be consumed at least partly in the formation of intermediate products with high energies which, by interaction with the original substance, are constantly renewed at the expense of the released energy. A simple case of such a chain reaction is afforded by the photochemical union of hydrogen and chlorine. The absorption of one quantum by a chlorine molecule produces, according to Einstein's law, an elementary act leading to the formation of two atoms of chlorine. Each chlorine atom then takes part in a cycle of changes represented by the equations

$$1.\quad Cl_2+h\nu=Cl+Cl;$$

$$2.\quad Cl+H_2=HCl+H;$$

$$3.\quad H+Cl_2=HCl+Cl;$$

and the chain of reaction thus proceeds until it is terminated by collision of a chlorine or hydrogen atom with an oxygen molecule, for example, or with the wall of the vessel.

The theoretical and experimental investigation of chain reactions has been the object of intensive study in recent years, and more than half the material in the book has been gathered in the last four years, a substantial part of it being done in the author's laboratory in Leningrad. The first part of the book deals with the general principles of the chain theory, together with the straightforward mathematics involved. The other three parts contain a detailed and critical analysis of the experimental data for nearly fifty reactions on the basis of the theory, and finally a summary of the main conclusions reached.

Such a survey by a leading authority is of very considerable value, particularly as the subject is one on which much work in various fields is now in progress. It is obviously impossible to specify the various reactions considered, but it may be mentioned that they include the classical hydrogen-chlorine reaction, the glow of phosphorus and the oxidation of hydrogen, carbon monoxide and hydrocarbons—all reactions which have been studied over a long period of years and still present obscure features. Prof. Semenoff's work is one which brings together in a very critical and instructive way a large mass of scattered information, and its appearance is an event which will be welcomed both by the specialists in this field and by the larger body of scientists who are desirous of finding out what progress has been made in recent years in the study of these very fundamental problems.

J. R. P.

Some Problems of Modern Meteorology. Pp. v + 170. (London: Royal Meteorological Society, 49 Cromwell Road, S.W. 7.) 3*s*. 6*d*.

This book consists of sixteen articles by fourteen authors, on some of the main problems of meteorology. These articles are reprinted, with some additional notes, from the *Quarterly Journal of the Royal Meteorological Society*, where they originally appeared during the period July 1930 to October 1934, under the editorship of Prof. D. Brunt, who contributes an introduction to the present volume. The subjects discussed are the origin of cyclonic depressions, radiation and absorption in the atmosphere, energy-transformations in the atmosphere (Brunt), weather-forecasting, rainfall (Douglas), origin of anticyclones, post-glacial climates and the forests of Europe (Brooks), the atmospheric circulation (Barlow), thunderstorms (Watson Watt), atmospheric turbulence (Davies and Sutton), seasonal forecasting (Normand), antarctic meteorology (Kidson), terrestrial magnetic variations (Goldie), meteorological acoustics (Tucker), condensation in the atmosphere (Bennett), and atmospheric ozone (Dobson and Meetham). The articles summarize the present position of these problems, and give numerous recent references to work upon them.

The collected re-issue of these articles should appeal not only to meteorologists, who should in general have read them in the *Quarterly Journal of the Royal Meteorological Society*, but also to a much wider circle of physicists who do not see that journal. The latter will, it is hoped, be impressed by the great body of meteorological knowledge which underlies these articles, and no less impressed and interested by the fundamental character of many of the chief unsolved problems of the subject.

The extremely moderate price of the book should be noted. S. C.

Thermostats and Temperature-Regulating Instruments, by ROOSEVELT GRIFFITHS, M.Sc. (London: Chas. Griffin and Co., Ltd., 1934.) 10*s*. net.

It is perfectly true that an instrument is a means to an end, and that the end should not be lost sight of in contemplating the means. Nevertheless among the devotees of experimental science one may find here and there a few favoured souls who take an artist's delight in an instrument for its own sake, who realize fully that the first duty of the instrument is to do its job, but whose interest slackens a little when they are asked to transfer their attention from the instrument to the job which it is doing. Whatever may be the bent of the author of this book there is no doubt that he is a lover of instruments. His descriptions of the various thermostatic devices are full, lucid and written *con amore*.

He has covered a wide range, and has described thermostats based on the temperature expansion of solids, of liquids and of gases, on the boiling-points of liquids, and on change of resistance with temperature. He has provided chapters dealing with the bimetallic and contact types of regulator, and with potentiometric and induction regulators. The last chapter of the book deals with low-temperature control. The volume is well produced and illustrated, and is compact in size and reasonable in price. It should prove a useful addition to the shelves of the experimenter.

A. F.

Principles of Mathematical Physics, by W. V. HOUSTON. Pp. xi+265. (London: McGraw-Hill Publishing Co., Ltd., 1934.) 21*s*. net.

This book is so condensed that it could be of little use except as a class-book where the lecturer proposes to cover practically the same course as that outlined in it. A feature is the large collection of problems, many of which we are more accustomed to seeing as theorems. Thus the student is given the problem of the brachistochrone as exercise 2 of chapter 5, and elsewhere is asked to find the field inside a spherical cavity and also inside long and short cylindrical cavities hollowed out in a block of magnetic material placed in a magnetic field. From the equation $E=mc^2$ he is asked to compute the mass of a moving electron, remembering that owing to the Lorentz transformation it becomes an ellipsoid of revolution. These examples suggest that the author is contemplating readers who will not be troubled by purely mathematical difficulties; yet apparently he envisages readers who need instruction in quite elementary differential equations, to which chapters 1 and 3 are devoted. Indeed, quite a considerable part of the book deals not with mathematical physics but with physical mathematics. In addition to the chapters just mentioned, there are others on the calculus of variations and on vector analysis, which really belong to the realm of mathematics. There is good reason for the inclusion of the one on variational methods, because there are few elementary treatments of this subject available. The chapter on vectors, again, is valuable for its last few sections, on the linear vector operator and on Gibbs's dyads; the relation to tensor calculus is shown, and use is made of these ideas elsewhere in the book, particularly in treating rigid dynamics by means of Euler's angles.

The part of the book devoted to true mathematical physics restricts itself to non-quantum physics; it is none the worse for this, since the main mathematical points which the student will need later in quantum theory have been rather ingeniously introduced in the treatment of the classical theory. The subjects dealt with are mechanics up to the Hamiltonian equations (but not elasticity or hydrodynamics), the first two laws of thermodynamics, statistical mechanics (14 valuable pages), electricity, and the restricted theory of relativity. A particularly good chapter is that on the theory of vibrating systems in general, where a digression is made to discuss the expansion of functions in series of orthogonal functions. On the other hand, it is a pity to see the author apologizing to the student for dealing with the mechanics of rigid bodies, and admitting that they do not exist in nature. Actually even a liquid is a rigid body, as far as the equations of mechanics are concerned, if its volume and shape are not changing.

The book is well printed and singularly free from misprints, except on p. 166 where the last few lines have no connection with the matter immediately preceding or following them.

J. H. A.

Relativity Physics, by W. H. McCrea, M.A., B.Sc., Ph.D., F.R.S.E. Pp. vii + 87. (London: Methuen and Co., Ltd., 1935.) 2*s*. 6*d*. net.

This short monograph gives an admirable account of those results of the special theory of relativity which find most frequent application in physics. There is no other book, as far as the knowledge of the reviewer goes, which contains in such an accessible and concise form the wealth of information which is provided here by Dr McCrea. After a brief account of the origin of the Lorentz transformation the author applies the special theory of relativity to numerous problems of kinematics, mechanics, optics, electromagnetic theory, atomic physics, thermodynamics, statistical mechanics and hydrodynamics. This is an amazing array of topics and every student of the theory of relativity will be glad of such a concise and simple summary of the principal results which have been obtained in these widely different fields by relativistic methods.

The only notable omission is the lack of any reference to the use which was made by L. H. Thomas of the Lorentz transformation in calculating the precession of a spinning electron moving through an electric field. This investigation has a considerable historic interest both in relation to the quantum theory of the anomalous Zeeman factor and in relation to relativistic rigid dynamics.

The mention of rigid dynamics suggests another small criticism which can be urged against this monograph. It deals exclusively with the successes of relativistic theory and makes no reference to its failures. Thus it does not refer to the many unsuccessful attempts to construct a relativistic dynamics of rigid or quasi-rigid bodies, and it does not refer to the, as yet, unresolved problem of the relativistic dynamics of a system of particles. The only reference to this topic is on p. 79, lines 4 and 5, where it is asserted that Hamiltonian equations of motion hold good in relativistic mechanics. There is a reference to the equations deduced on p. 28. However, these equations refer to a single particle only and all attempts to construct a Hamiltonian function for a system of several particles have so far met with complete failure. The difficulty is probably due to the inapplicability of the special theory of relativity to any systems except isolated particles. Thus Eddington has recently asserted the need for reintroducing the concept of absolute simultaneity when discussing the properties of a complex system.

However, these detailed criticisms are entirely of a minor character and do not detract in any way from the great value of this little book as a most valuable summary of standard results.

G. T.

Solid Geometry, by L. Lines. Pp. xx + 292. (London: Macmillan and Co., 1935.) 6*s*.

The disadvantage in allowing one who is not a practising teacher of mathematics to review this book is that the reviewer is not sure where to apportion his praise. The book is intended primarily for the use of pupils preparing for one of the Higher School Certificate examinations, and thus gives a course in solid geometry which would frequently be taken by those not destined to become mathematical specialists. From this point of view, the advance on the methods of twenty years ago is enormous and somebody should be highly congratulated. If Mr Lines is leading the way, let him accept the felicitations; if he merely follows others, then we must regret that the pioneers have not been more widely acclaimed.

The solid geometry dealt with is the Euclidean geometry of the point, line and plane, extended so as to make use, where advisable, of algebraic and trigonometric methods. It is not 'experimental' geometry, and on the other hand it does not overlook the value of suitably drawn diagrams. The striking differences from the older books lie in the subject matter, and the use made of modern ideas in notation. To take the latter first, we can illustrate it with part of the rider "If three planes intersect, two by two, show that their lines of intersection are either concurrent or parallel".

> Let the three planes be α, β, γ. The straight lines $\gamma\alpha$ and $\alpha\beta$ lie in the same plane α; $\therefore$ they either intersect or are parallel.
>
> *If they intersect*, their common point lies on all three planes; i.e. it is $\alpha\beta\gamma$;
>
> $\therefore$ it lies on $\beta\gamma$
>
> $\therefore$ $\beta\gamma$, $\gamma\alpha$, $\alpha\beta$ are concurrent."

The way the notation assists the argument is well brought out even to a young pupil.

As regards the subject matter, after the more usual geometry and mensuration we come to a detailed treatment of polyhedra, including the star polyhedra, and it is pointed out that Euler's theorem breaks down for certain of the latter. These chapters are followed by extremely good treatments of subjects suggested by physics—space lattices and their classification, and sphere-packs—and the book is brought to a close with a chapter on patterns and crystals. Perhaps the only thing to regret is the absence of some elementary introduction to geometry on the surface of a sphere.

A question arises whether the average young student will find the nomenclature too much strain on the memory. The first page of the excellent index includes *anorthic*, *bisphenoid*, *Bravais*, *clinographic*, *reguli*, *snub* (cube), *cuboctahedron*, *dyakis*, *stellated* (dodecahedron) and *enantiomorphous*, all of them words which would be new to most pupils.

J. H. A.

Cambridge Tracts in Mathematics and Mathematical Physics, No. 32. *Generalized Hypergeometric Series*, by W. N. Bailey. Pp. 108. (London: The Cambridge University Press.) 6*s*. 6*d*. net.

The latest of the *Cambridge Tracts in Mathematics and Mathematical Physics* is devoted to generalized hypergeometric series. This is a subject in which most of the developments of the last few years have been made by the author, Dr W. N. Bailey. He has skilfully arranged his material so as to display the variety in the methods which are available for obtaining the numerous formulae which have been discovered.

A hypergeometric series is a generalization of the binomial expansion. Binomial coefficients can be found in succession by multiplying by fractions, the numerators and denominators of the fractions being increased successively by unity. Multiply each term by two such fractions and the hypergeometric series is obtained. Multiply by three or more such fractions and the resulting expansion is a generalized hypergeometric series. It is well known that certain hypergeometric series are related to each other in simple ways. In some cases the series can be summed, just as a binomial expansion can be summed. The more general theory is concerned with the discovery of relations between generalized hypergeometric series and with their summation, which is usually in terms of gamma functions.

It has been found possible to classify the generalized series; the most interesting type is the "well-poised" series in which each of the fractions used for multiplication at any stage of the generation of the series has the same value for the sum of numerator and denominator. Bailey has actually found formulae relating to well-poised series of the ninth order, i.e. series in which there are no fewer than nine fractions to multiply by at each stage.

The subject has many points of contact with the theory of the functions used in mathematical physics and, although the more elaborate formulae in the tract have not yet found any application, it may prove useful to physicists to have such formulae readily available. There is one application which has not yet been made explicitly, though the cell for it was announced in 1858 when Cayley published a short article in the *Philosophical Magazine* and enunciated a proposition which he had not been able to prove but which he said he had discovered in discussing certain relations in planetary theory. A proof of Cayley's theorem was published by Orr in 1899 and it is now known how the theorem fits into the theory of generalized hypergeometric series, but it is not known whether there is an important application in planetary theory. It is to be hoped that some expert in that subject will look into the question.

F. J. W. W.

Numerical Studies in Differential Equations. Vol. **1**, by H. LEVY and E. A. BAGGOTT. Pp. viii + 238. (London: C. A. Watts and Co.) 12*s.* 6*d.*

All who were present at the fascinating lecture given by Prof. Levy to the London Mathematical Society on May 19, 1932, have been looking forward ever since to the appearance of his promised book on the numerical treatment of differential equations. The present instalment, volume **1**, tends rather to whet the appetite for what is to come than to slake the thirst. Here we have an expanded treatment of the purely graphical methods and of the usual numerical processes of the type devised by Runge. Picard's method for obtaining a series solution (not necessarily in integral powers of the independent variable) is also included, and is made to appear much more attractive than most writers have made it.

The familiar methods of integrating ordinary differential equations numerically practically restrict themselves to the case when the boundary conditions consist in the assignment of definite values to the dependent variable and all but the highest of its differential coefficients, at a point which may be taken as origin. This book does not shirk the more difficult problems which arise when boundary conditions are assigned at two or more points in the range of integration, and it is in this respect that it will be most valuable.

The treatment is very clear throughout. As befits the subject, processes and methods are frequently expounded not by giving general investigations but by applying them to specific numerical problems. In addition, an adequate selection of examples for practice is included.

The last chapter contains a variety of miscellaneous theorems which do not strictly belong to the subject of numerical treatment, though their applicability to that subject is illustrated—in striking fashion on occasion. Here we are reminded by meeting the characteristic values of a parameter, that the second volume is to deal systematically with partial differential equations. This will be a difficult task, but the manner in which volume **1** has been carried out suggests that the right authors have been found for the work.

J. H. A.

Mathematical Tables. Vol. **3**.* (*Minimum Decompositions into Fifth Powers.*) By L. E. DICKSON. (London: Office of the British Association, 1933.) 10*s.*

The theorem that any integer can be expressed as the sum of not more than four squares has long been known, but there is no corresponding theorem for powers higher than the square. In the case of the fifth powers, for example, it is evident that 31, being less than 2^5, cannot be decomposed into fewer than 31 such terms, whilst 223 (which is less than 3^5) requires 37, viz. $(6 \times 2^5 + 31 \times 1^5)$.

The present table shows the actual minimum decompositions of all numbers up to 150,000 and the minimum number of fifth powers to be added together to make any number up to 300,000. Thus the entry

87701	1 0 5 1 6 0 1

---

* In the review of volume **2** of this publication, *Proc. phys. Soc.* **45**, 133 (1934), "volume **3**" was printed in error for "volume **2**."

means that $87701 = 2^5 + 0 \times 3^5 + 5 \times 4^5 + 1 \times 5^5 + 6 \times 6^5 + 0 \times 7^5 + 1 \times 8^5$. Since it is evident that when the minimum decomposition of N is known, a decomposition of $(N+1)$ can be obtained merely by adding 1^5, space can be saved by omitting the entries for those numbers for which this decomposition is the minimum one. This has been done here, with the result that the numbers up to 10,000 occupy only about 1050 entries, and are set out in four and a half pages. After 150,000 such an entry as $\boxed{179379}\boxed{12}$ means that 12 fifth powers will suffice for this number. The actual decomposition is quite easy to find: subtracting 11^5 or 161051 from 179379 gives 18328, which (from the earlier part of the table) has more than 11 decompositions. Hence the first term is not 11^5. Next taking 10^5 as a trial value, we find from the table that 79379 has too many decompositions. The next trial, with 2×9^5, leaves a remainder $61281 = 3^5 + 4 \times 4^5 + 3 \times 6^5 + 2 \times 7^5$, making the correct total of 12 fifth powers for the original number.

The number of fifth powers to be added together to make a given integer seems on the whole to decrease with the size of the number. No number in the table requires more than 37 terms, and this occurs only for $N = 223$. After this, no number requires more than 31 fifth powers; after 30,000 no number up to 300,000 requires more than 19; and between 191,263 and 4,037,824 it has been proved that 15 terms will always suffice. It is also interesting to note that the results given here have been used by Dickson to prove that 37 gives an upper limit for numbers right up to 10^{483}.

The book is well printed (it is reproduced by the replica process from typescript, to reduce the risk of error), but it might have been convenient if a table of fifth powers had been printed as an appendix.

J. H. A.

THE PROCEEDINGS OF
THE PHYSICAL SOCIETY

VOL. 47, PART 4 July 1, 1935 NO. 261

534.22

THE EFFECT OF PRESSURE ON SUPERSONIC DISPERSION IN GASES

BY W. RAILSTON, M.SC. AND E. G. RICHARDSON, B.A., D.SC., PH.D. (Armstrong College, Newcastle-on-Tyne)

Received February 14, 1935. *Read April* 5, 1935.

ABSTRACT. Measurements of wave-lengths have been made by interferometer and hot-wire methods, and of absorption by hot-wire methods alone, for supersonic radiation at frequencies between 40 and 2000 kc./sec. in carbon dioxide, nitrous oxide and sulphur dioxide at various pressures up to two atmospheres. For the hot-wire method, a circuit which gives a linear relation between amplitude or particle velocity and response at constant frequency is described. The velocity-measurements in the former two gases may be reduced to a common curve by plotting them against the parameter {frequency/pressure}; although the observed rise in velocity in the region where this parameter lies between 100 and 1000 is in accordance with the relaxation-time theory, the decrease in velocity at lower and higher values is not. In sulphur dioxide the rise of velocity is in the neighbourhood of 4000. The absorption in all three gases rises sharply at pressures below 400 mm. of mercury.

§ 1. INTRODUCTION

It was formerly supposed that owing to the interdependence of pressure and density the velocity of sound in a gas was immune from variations due to a change of pressure in the gas. Pierce⁽¹⁾ and others having shown that the velocity rises above the normal in certain gases at frequencies in the supersonic region, it becomes of interest to observe what influence, if any, pressure has on this phenomenon. As regards absorption, the well-known Kirchhoff formula gives for the absorption coefficient

$$\alpha = \frac{n^2}{2V^3\rho}\left(\frac{4}{3}\eta + \frac{\gamma-1}{\gamma}\frac{\kappa}{s}\right)$$

or

$$\alpha\lambda^2 = \frac{1}{2V\rho}\left(\frac{4}{3}\eta + \frac{\gamma-1}{\gamma}\frac{\kappa}{s}\right) \quad \text{......(1)},$$

α

where V is the velocity, n the frequency and λ the wave-length of the sound, ρ the density, η the viscosity, κ the conductivity, s the specific heat, and γ the ratio of specific heats of the gas. The energy-absorption coefficient would be double these values. This formula would indicate that as the pressure is reduced the absorption should increase if the frequency remains constant, but again, coefficients many times greater than the Kirchhoff values have been measured in certain gases. Although explanations of these anomalies have been put forward and will be considered later, it seems desirable that the effects of both pressure and temperature on these phe-

V, n, λ, ρ
η, κ, s, γ

nomena should be investigated, if the experiments on dispersion of this kind are to be put on a theoretical basis. The object of the present paper is to report experimental work on the former of these effects, viz., that of pressure. Previous work in this direction has not been extensive. The earlier measurements of Koch(2) on air compressed from 1 to 200 atmospheres and the more recent ones of Spakowski* on carbon dioxide from 1 to 85 atmospheres relate to a single frequency within the audible region, and show a steady decrease of velocity with increasing pressure at constant temperature. During the progress of our work Richards and Reid(4) published a few results of supersonic velocities at the two frequencies 94 and 450 kc./sec. and various pressures.

Apparently the dependence of absorption on pressure has never before been investigated.

§ 2. APPARATUS

The majority of the velocity-measurements were made by the Pierce method in which the reaction upon the driving circuit of a piezoelectric quartz oscillator due to the slow displacement of a reflecting surface in a direction perpendicular to the crystal face was used to measure the wave-length, the frequency being measured at the same time on a wave-meter. Other measurements, both of wave-length and absorption, were made by the method already used by one of us (E.G.R.) at atmospheric pressure(5), viz. a method in which the reflector and source are kept still, while a hot-wire anemometer is traversed across the space between them.

One of the difficulties which we have experienced in the use of the Pierce method is that the varying reaction which the quartz oscillator experiences as the reflector is moved back tends to make the oscillator uncertain in action, so that it is difficult to find the peaks of current unless the maintaining circuit is either very powerful or very stable. The use of high-power valves, however, brings other difficulties in its train, the chief of which is the production of unwanted overtones in the crystal with consequent subsidiary peaks. The maintenance circuit finally adopted is shown on the right of figure 1. When the switch S_1 is closed and the oscillator circuit LC is tuned, the oscillations of the crystal Q are maintained by the high-impedance valve (LS 5) in the fashion of the usual Pierce circuit, and the peaks shown by the slight movement of the milliammeter needle MA as the reflector is moved. But when the switch S_2 is closed, the oscillatory part of the plate current in the first valve is coupled through the h.-f. nickel-cored transformer with reaction, T, to the second valve LP 2 in such a way that the galvanometer G shows quite large peaks. The sensitivity of the galvanometer may be varied by the shunt R_1 and the peaks and troughs in the readings made to lie evenly about the zero of the scale by adjustment of the resistance R_2 connected in series with the backing battery B. This circuit has proved to be most satisfactory in use, and we confidently recommend it to those who require sufficient power with stability to operate the Pierce method. The choke and transformer should be surrounded by earthed screens.

---

* See reference (3). This author omits to record the frequency! From the description of the apparatus we infer that it was in the audible region.

To the left of figure 1 is shown the hot-wire amplifier circuit. The hot wire experiences a cooling which is proportional, at a fixed frequency, to the mean velocity in the alternating air current engendered by the supersonic waves; unfortunately the relation between velocity and change of resistance is not a linear one. King (6) has given two formulae for the loss of heat, H, experienced by a wire of diameter d in a fluid of specific heat s, thermal conductivity κ and density ρ, viz.

H

d, s, κ, ρ

$$H = 2\pi\kappa\theta \log^{-1}(2b/d), \text{ when } Vd/\nu < 10 \qquad \text{......(2)},$$

$$H = \{\kappa - (2\pi\kappa s\rho Vd)^{\frac{1}{2}}\}\,\theta, \text{ when } Vd/\nu > 10 \qquad \text{......(3)},$$

in which θ is the excess temperature, ν the kinematic viscosity of the fluid,

θ, ν

$$b = \kappa e^{1-\gamma}(s\rho V)^{-1},$$

b

and γ is the ratio of specific heats.

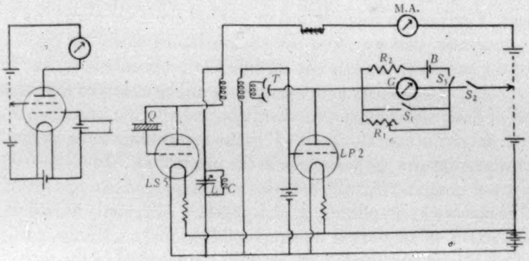


Figure 1. Hot-wire and oscillator circuits.

In the present measurements we are dealing with the first region, and since the resistance-change is given in the form

$$R = R_0(1 + a\theta), \text{ while } H = i^2R,$$

the relation between the resistance and the velocity is, *ceteris paribus*, of the type

$$R \log V = \text{constant}$$

or

$$V = \exp(1/R) + \text{constant}.$$

In order to amplify the small change in resistance of the hot wire when it is exposed to supersonic radiation, and at the same time to get a linear relation between the decrease of resistance and increase of velocity, we have adopted a device which Luneau (7) has employed on the ordinary hot-wire anemometer for aeronautical research. This consists in putting the wire with its heating-battery in the grid-filament connection of a valve, while a galvanometer is placed in the plate-filament connection, figure 1. Luneau (7) was concerned with high air velocities and therefore used a valve whose characteristic in the working region was such as to turn a relation of the type

$$\delta R \propto \delta(V^{\frac{1}{2}})$$

into one as nearly as possible like

$$\delta I \propto \delta V,$$

1 where I is the current in the galvanometer. In this instance we use a variable-mu valve (*VS* 24) so that the amplification is proportional to the logarithm of the grid resistance. We have calibrated the hot wire, connected in this way, with small steady air-velocities and found that a linear relation exists between I and V over the range required. Further, the grid-bias battery may be adjusted to such a value (-2 V. in this case, with an anode and screen potential of 75 V.) that no anode current flows when the hot wire is in a still gas. Both the resistance of the wire in still air and its sensitivity to velocity varied with the pressure. The natural convection current is a function of the density and in fact increases with the pressure of the gas, so that the resistance in still gas decreases as the pressure goes up. The coefficient of variation with pressure is not constant for all gases; Petavel(8), for example, found that whereas in air the natural convection current varied approximately as the square root of the pressure, in carbon dioxide the proportionality was nearly linear. Certainly in this gas and in nitrous oxide we found, by comparison with the manometer, that we could use the resistance when the gas was still as a measure of the pressure. In sulphur dioxide the proportionality was less certain. As regards change of sensitivity with pressure, it will be observed that in the formula for the loss of heat, although the conductivity and specific heat change little with pressure, the density occurs along with V in the logarithmic term, so that in a given gas the hot wire measures the variations in the product ρV. In conformity with this deduction it was observed that the hot wire was more sensitive to a given change of V, i.e., of displacement-amplitude, as the pressure increased; indeed at pressures below about 10 cm. of mercury it was impossible to make readings, partly owing to this lack of sensitivity and partly because with the removal of the gaseous load the crystal refused to oscillate with sufficient amplitude.

The vessels used were of steel or glass, 10 cm. in diameter. A steel cylinder was used for the Pierce measurements, each end being closed with a plate of thick brass. Part of the thickness of the brass was turned down to make a friction fit over the steel tube, leaving a projecting lip on to which molten solder was poured to ensure gas-tightness. The reflector used was an ebonite disc set accurately perpendicular to a steel screw of pitch 0·635 mm. and diameter 6 mm. At the point of insertion of the screw in the end cap a gland holding a mixture of vacuum grease and paraffin wax was added. Sparking-plugs were found to be the most suitable leads to the electrodes which were of staybrite steel.

The glass cylinder was used for the hot-wire measurements. In this case the apparatus was similar to that already used at atmospheric pressure(5), except that greater care was taken in sealing. The apparatus could be exhausted to 3 mm. of mercury and was filled by repeated flushing and evacuation.

The gases used were obtained directly from cylinders after being passed over calcium chloride. Pressures up to two atmospheres were used, the pressure being read on a mercury manometer connected to a tap on the cylinder. As has been explained above, the pressure could also be checked by reading the resistance of the wire when this was heated by the standard current of 0·2 A. The temperature was also taken at each set of readings. The frequency was measured on a calibrated wave-meter with the quartz crystals *in situ* during the experiments, since a small variation of frequency with the air gap in the holder has been observed.

§ 3. RESULTS

The mode of operation of the apparatus was as follows. The driving circuit being turned on, the variable condenser was set for the optimum maintenance conditions of the piezoelectric oscillator and the change-over switches operated to set the galvanometer into operation at suitable sensitivity and at the zero mark on the scale. When the hot wire was used, the current through it and the backing battery of its galvanometer were adjusted before the quartz was started into oscillation. The reflector was then moved back step by step through some dozen half-wave-lengths, while readings of one or both of the galvanometers were made. The reflector was then brought back while a similar series of readings were obtained. Readings of the pressure, temperature and frequency were then made, and the pressure was changed.

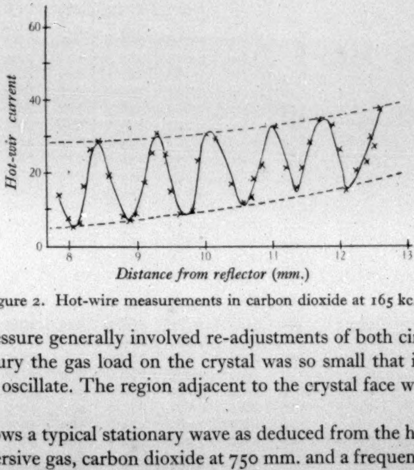


Figure 2. Hot-wire measurements in carbon dioxide at 165 kc./sec.

A change of pressure generally involved re-adjustments of both circuits, and below 10 cm. of mercury the gas load on the crystal was so small that it was difficult to get the latter to oscillate. The region adjacent to the crystal face was avoided in the measurements.

Figure 2 shows a typical stationary wave as deduced from the hot-wire measurements in a dispersive gas, carbon dioxide at 750 mm. and a frequency of 165 kc./sec. The supersonic amplitude ξ at any point x on the curve is given by ξ, x

$$|\xi|^2 = 2A^2[\cosh\{2\alpha(x-l)\} - \cos\{2\beta(x-l)\}],$$

where α is the absorption coefficient, $\beta = 2\pi/\text{wave-length}$, l is the distance of the reflector from the source, and A is a constant. The amplitude varies between β, l

$$2A\cosh\{\alpha(x-l)\}$$

and

$$2A\sinh\{\alpha(x-l)\},$$

so that these expressions define the loci of the peaks and troughs (broken lines) and allow α to be calculated (5).

In the tables which follow the results of the wave-length measurements in the three gases at two frequencies are given as examples. The pressure in mm. of mercury appears in the first column; it ranged over nearly two atmospheres in carbon dioxide and nitrous oxide, but we had no ready means of compressing the sulphur dioxide, and therefore results up to atmospheric pressure only are given for this gas. The second column gives the temperature in degrees centigrade, the third the measured wave-length in millimetres and the fourth the velocity in metres per sec. reduced to 0° C.

Table 1

Carbon dioxide				Nitrous oxide				Sulphur dioxide			
p	θ	λ	V_0	p	θ	λ	V_0	p	θ	λ	V_0
Frequency 98 kc./sec.											
99	21·0	2·859	270·0	164	21·0	2·824	266·7	106	20·0	2·208	208·9
152	20·5	2·820	265·3	217	21·5	2·809	265·0	153	20·0	2·206	208·7
207	21·0	2·810	265·2	253	21·5	2·812	265·3	217	20·5	2·207	208·6
305	21·0	2·810	265·2	307	21·0	2·809	265·2	274	21·0	2·202	208·0
353	21·0	2·798	264·2	368	21·0	2·798	264·2	338	21·0	2·208	208·5
457	21·5	2·790	263·1	479	20·5	2·782	262·9	346	21·0	2·205	208·3
545	21·0	2·775	262·0	572	21·0	2·775	262·0	392	20·5	2·209	208·6
608	21·0	2·765	261·1	634	21·0	2·769	261·5	451	20·5	2·205	208·4
716	21·5	2·757	260·1	732	21·0	2·721	257·0	507	20·0	2·202	208·3
754	21·5	2·739	258·4	754	21·0	2·712	256·3	563	20·0	2·202	208·3
802	21·5	2·742	258·7	819	20·5	2·703	255·6	629	20·5	2·201	208·1
924	21·0	2·733	258·1	875	20·5	2·701	255·4	670	20·5	2·205	208·4
1016	20·5	2·723	257·2	921	20·5	2·694	254·8	717	21·0	2·205	208·3
1113	21·0	2·723	257·2	1063	21·0	2·699	254·8	758	21·0	2·206	208·4
1200	21·0	2·723	257·2	1114	21·0	2·693	254·3	—	—	—	—
1268	21·0	2·721	257·0	1238	20·5	2·684	253·8	—	—	—	—
1303	21·5	2·717	256·4	1291	20·5	2·683	253·6	—	—	—	—
1352	21·5	2·713	256·0	1354	21·0	2·682	253·4	—	—	—	—
1456	21·0	2·708	255·8	1416	21·0	2·680	253·2	—	—	—	—
1552	21·0	2·703	255·3	1537	20·5	2·665	252·1	—	—	—	—
Frequency 1000 kc./sec.											
252	19·5	0·279	269·5	264	20·5	0·278	268·5	209	20·0	0·219	211·0
307	19·5	0·279	269·5	317	20·5	0·279	269·0	253	20·5	0·218	210·0
354	19·0	0·279	269·5	351	21·0	0·279	269·0	301	21·0	0·217	209·0
402	19·0	0·279	269·5	392	21·0	0·279	269·0	356	21·0	0·217	209·0
458	19·0	0·280	271·0	476	20·5	0·280	270·0	422	21·0	0·217	209·0
522	19·5	0·281	271·5	558	20·5	0·281	271·0	479	20·5	0·217	209·5
573	19·5	0·281	271·5	597	20·5	0·281	271·0	564	20·5	0·217	209·5
634	19·0	0·280	271·0	645	20·0	0·281	271·5	598	20·5	0·216	208·5
681	19·0	0·280	271·0	682	20·0	0·281	271·5	632	21·0	0·216	208·5
785	19·5	0·280	270·5	746	20·0	0·281	271·5	697	21·0	0·216	208·5
837	19·5	0·280	270·5	831	19·5	0·279	270·0	740	21·0	0·216	208·5
886	20·0	0·280	270·0	874	20·0	0·279	269·5	—	—	—	—
949	20·0	0·279	269·5	930	20·0	0·279	269·5	—	—	—	—
1018	20·0	0·280	270·0	983	20·5	0·279	269·0	—	—	—	—
1063	20·0	0·280	270·0	1068	21·0	0·279	269·0	—	—	—	—
1127	20·5	0·279	269·0	1127	21·0	0·280	269·5	—	—	—	—
1232	20·0	0·280	270·0	1193	21·0	0·279	269·0	—	—	—	—
1294	20·0	0·279	269·5	1261	20·5	0·279	269·0	—	—	—	—
1358	20·0	0·279	269·5	1349	20·5	0·279	269·0	—	—	—	—

On figures 3 and 4 (not drawn to the same scale) the results are grouped in the form of reduced velocity against pressure for all the frequencies used, in carbon dioxide and sulphur dioxide respectively. The results in nitrous oxide were so similar to those in carbon dioxide as scarcely to warrant recording on a separate graph, but in figure 5 we have certain of the velocity results in these two gases

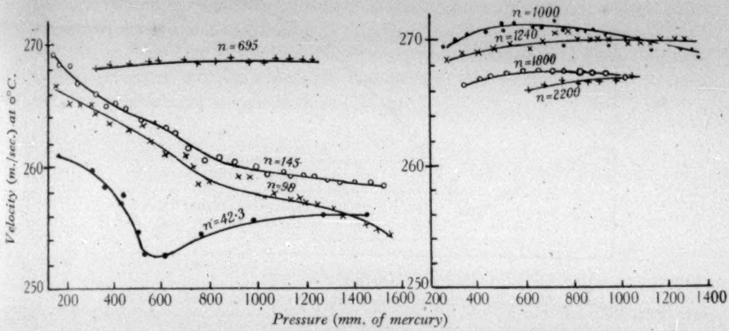


Figure 3. Reduced velocities in carbon dioxide. The figures on the curves give the frequency in kc./sec.

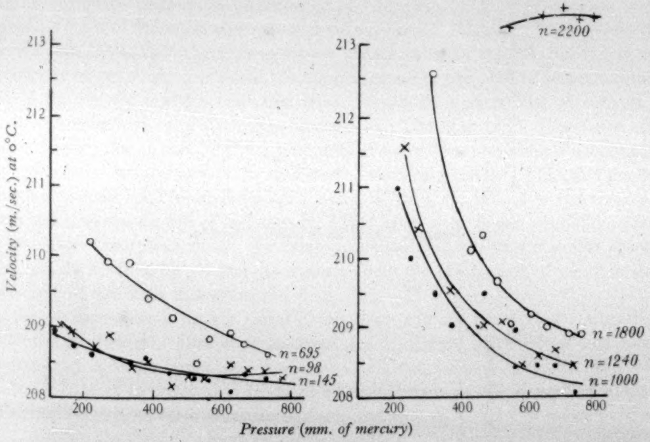


Figure 4. Reduced velocities in sulphur dioxide. The figures on the curves give the frequency in kc./sec.

recorded together against the parameter log (n/p), n being in c./sec. and p in mm. of mercury. The dispersion is apparent as a rise of velocity in the neighbourhood of $n/p = 100$, with a more gradual fall on each side (9). In sulphur dioxide, figure 6, this effect is only beginning to occur at the maximum values of n/p attained. The

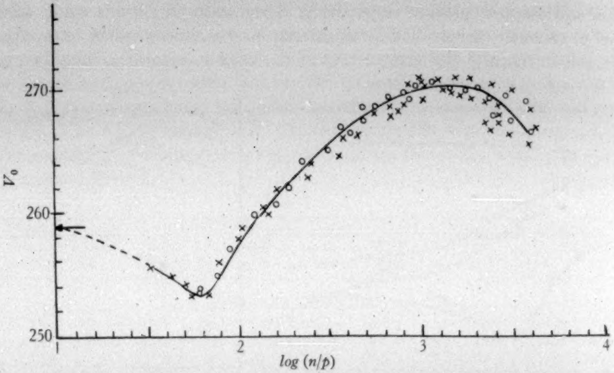


Figure 5. Velocities in carbon dioxide (crosses) and nitrous oxide (circles).

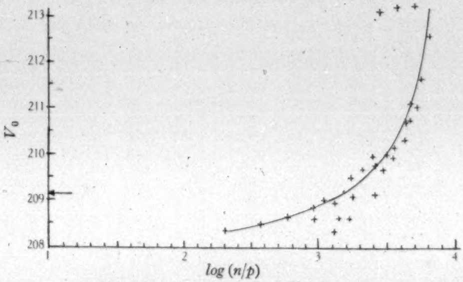


Figure 6. Velocities in sulphur dioxide.

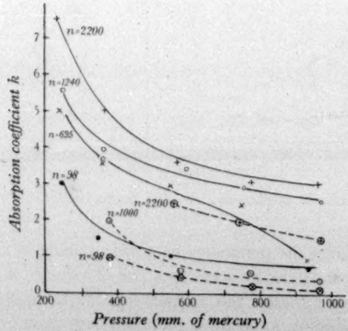


Figure 7. Absorption in carbon dioxide, nitrous oxide and sulphur dioxide. The observation points for the latter gas are enclosed in circles and joined by broken lines.

arrows on these two figures indicate the value of V_0 at low frequencies. Finally on figure 7 we have the absorption coefficient k plotted against pressure at four frequencies. In each case the absorption rises as the frequency is increased or the pressure reduced, but plotting against n/p does not reduce the results to a single curve so successfully as it did in the case of the velocities. It must be remembered however that α cannot be determined with the same accuracy as λ by a hot-wire method, or indeed any other method. In table 2 are shown the values of $k\lambda^2$ from equation 1, which are independent of pressure.

Table 2

Pressure (mm. mercury)	200	360	560	760	1000
$k\lambda^2 \times 10^6$ (carbon dioxide)	7·5	4·0	2·7	2·0	1·5
$k\lambda^2 \times 10^6$ (sulphur dioxide)	5·5	3·7	2·4	1·5	1·1

§4. DISCUSSION OF RESULTS

Two theories have been advanced to account for the anomalous dispersion and absorption of supersonics in gases such as carbon dioxide. Either we may suppose that there is a lag in the transformation of one type of molecular energy into another, e.g. of energy of translation into energy of vibration, or we may suppose that a modified form of resonant absorption is taking place.* Kneser (10) has shown that on the first hypothesis there should be a steady value of the velocity below that frequency at which the period of vibration is equal to the time of relaxation or of lag in the energy transfer, and that there should be a sharp rise of velocity to another maintained value at higher frequencies. The absorption also, when $\alpha\lambda$ is plotted against log n, should give a symmetrical curve. With resonant absorption, on the other hand, assuming that the resonance is not sharp but broad like that due to a number of resonance frequencies spread over a band, we should as the frequency rises get a moderate fall of velocity on either side of a steeper rise and again a symmetrical absorption curve. We must remember that in either case this absorption may be clouded by viscous dissipation, about the mechanism of which at such high frequencies little is known.

As we have shown, figures 5 and 6, the dispersion is a function of the quotient {frequency/pressure}. This is explicable on a resonance theory since one would expect any molecular dimension or spacing which was responsible for the dispersion to be diminished in proportion as the pressure increases, so that if resonance occurs at a frequency n_1 under a pressure p_1 it will occur at n_2 under a pressure p_2 where $n_1/p_1 = n_2/p_2$. On the other hand Richards and Reid (4), who are the first to point out the efficacy of plotting the velocity against this parameter, though on the evidence of a more restricted range of experiment, suggest an explanation on the relaxation-time theory. Since the latter involves the *probability* of the transition from translational to vibrational energy, they assume this probability to be proportional to the pressure as a basis for an explanation of the results.

---

* Cf. reference (5) and the bibliography given therein.

On the absorption results it is less easy to compare the effect of pressure with theory since the issue is confused by the viscous absorption, which as equation (1) shows increases as the pressure diminishes. Apart from this there is a correspondence between absorption curves plotted against {frequency/pressure}, although the agreement is not so good as that obtained from the velocity curves plotted in this fashion. In agreement with Groszmann$^{(11)}$ we find absorption coefficients of rather higher order than those postulated by Kneser. The crux of the dispersion occurs at a value of n/p equal to 100 in carbon dioxide and nitrous oxide, figure 5, and beyond $n/p = 6000$ in sulphur dioxide, in which gas the rise of velocity is slow but is becoming rather more rapid at the limit of our n/p range, figure 6.

All that we would venture to state at present is that the relaxation theory in the form in which it is propounded by its advocates is insufficient to explain our observations. Before theorizing further it is desirable to make a comparison with the dispersion shown by a medium having resonators of macroscopic dimensions, such as a suspension having a known density and a known range of particle-size. Experimental work on such systems is now being initiated in this laboratory.

§5. ACKNOWLEDGMENTS

We wish to thank the Dixon Fund Committee of the University of London and the Research Committee of Armstrong College for Grants enabling the oscillators to be purchased and Professor W. E. Curtis for placing the facilities of the Physics Department at our disposal.

REFERENCES

(1) Pierce. *Proc. Amer. Acad. Arts Sci.* **60**, 271 (1925).
(2) Koch. *Ann. Phys.*, Lpz., **26**, 531 (1908).
(3) Spakowski. *C.R. Acad. Sci.*, Leningrad, **3**, 31 (1934).
(4) Richards and Reid. *J. Chem. Phys.* **2**, 198 (1934).
(5) Richardson. *Proc. roy. Soc.* **146**, 56 (1934).
(6) King. *Philos. Trans.* **214**, 373 (1914).
(7) Luneau. *Aeronautique*, **15**, 232 (1933).
(8) Petavel. *Philos. Trans.* **197**, 229 (1901).
(9) Penman. Page 543 of this volume.
(10) Kneser. *Ann. Phys.*, Lpz., **11**, 777 (1931).
(11) Groszmann. *Ann. Phys.*, Lpz., **13**, 681 (1932).

DISCUSSION

See page 777 for discussion.

534.22

THE EFFECT OF TEMPERATURE ON SUPERSONIC DISPERSION IN GASES

By H. L. PENMAN, M.Sc., A.Inst.P., Armstrong College,
Newcastle-upon-Tyne

Communicated by Dr E. G. Richardson, February 14, 1935. *Read April* 5, 1935.

ABSTRACT. Measurements of wave-lengths of supersonic radiation at frequencies between 40 and 140 kilocycles per sec. have been made in carbon dioxide, nitrous oxide and sulphur dioxide at various temperatures from room-temperature up to 200° C. The velocities at constant density, i.e. reduced to 0° C., have been calculated. When these are plotted against temperature, supersonic dispersion is shown by a sharp fall of velocity in carbon dioxide at a temperature which increases with the frequency of the source. The significance of these results in the light of the theories of supersonic dispersion which have been propounded is then discussed.

§ 1. INTRODUCTION

THE interpretation of supersonic dispersion in gases on the basis of a relaxation time as outlined by Kneser⁽¹⁾ has been extended in a recent series of papers by Richards and Reid⁽²⁾ to cover the effects of pressure and temperature. The theoretical treatment and practical results for the pressure effect are discussed by Railston and Richardson⁽³⁾. The present paper is a report on measurements made in a number of gases at various temperatures.

There is as yet no established theoretical basis for the observed temperature effects. Kneser postulated a temperature effect of zero order, that is to say, one independent of the change in velocity which results from a density-change. The early results obtained by Richards and Reid indicated that at high frequencies there was a variation of velocity with temperature and they suggested that it might be accounted for by introducing the idea of an activation energy of collision into theoretical considerations. This suggestion implies that only molecules having higher kinetic energies can cause excitation and de-excitation to and from higher vibrational levels. It leads to the conclusion that in the dispersive region the velocity of sound in a gas, being greater than the velocity in the audio-frequency range at room-temperature, will gradually approach the audio-frequency value as the temperature is increased.

The experimental data available are few. Apart from Richards and Reid's results already mentioned, further results were obtained by the same authors for sulphur dioxide and carbon disulphide, and in neither case did they fit the theoretical curves. Sheratt and Griffiths⁽⁴⁾ have published results of measurements in carbon monoxide made at frequencies of 7·9 and 27·4 kc./sec. in the temperature range 1000° to 1800° C. The difference between the velocities at these two frequencies increases as the temperature rises, the change thus occurring in the direction opposite to that predicted by the activation theory.

The present paper gives an account of measurements made on carbon dioxide at a number of frequencies in the dispersive region and also includes preliminary results for sulphur dioxide and nitrous oxide.

§ 2. APPARATUS

Quartz crystals piezoelectrically driven have been used as sources of radiation. The electrical driving-circuit was essentially that due to Pierce(5). A PX 25 valve was used with an anode voltage of 200, the crystal being connected between plate and grid. A small choke between the plate and the crystal-electrode was found to be effective in eliminating high-frequency harmonics and also in reducing the effects of hand-capacity. Measurements of wave-length were made in the standard way with an acoustic interferometer, a sensitive, suitably backed microammeter being included in the plate-filament circuit. Frequencies were determined at room-temperature with a wave-meter.

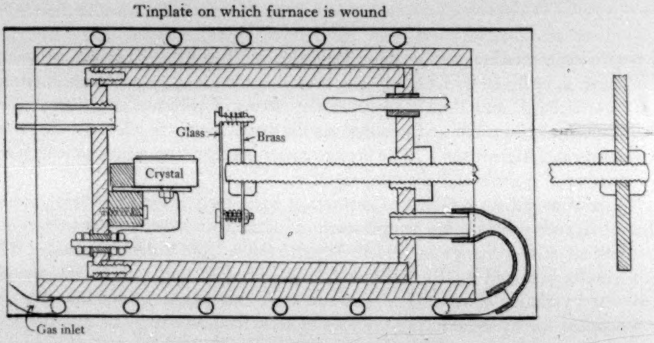


Figure 1. Diagram of apparatus.

Interferometer design. Two tubular containers for the gases have been used. The larger one was designed to accommodate the largest crystal employed. Its size prevented adequate temperature-control and occasionally there were doubts about its gas-tightness. The smaller tube substituted to overcome these objections is shown diagrammatically in figure 1.

The inner tube is of iron, $5\frac{1}{2}$ in. long and $2\frac{1}{2}$ in. in diameter. One end is permanently closed and carries a thermometer, a gas tap and a long brass screw on which the reflector is mounted. The other end is detachable and on it is mounted the crystal-holder, while through it pass a gas tap and two insulated terminals by means of which connexion is made to the crystal electrodes. This end is bolted on to the tube, a rubber gasket ensuring a gas-tight fit.

The crystal-holder is as simple as possible. The lower electrode, which is of thick brass with a plane clean surface, is merely a platform normal to the plane detachable end of the tube. The upper electrode is of thin aluminium resting lightly on the top of the crystal, connexion to the terminal being made through a very fine piece of fuse wire so that no constraint is imposed upon it. Two cork guides orient the crystal so that it vibrates in the direction of the axis of the tube, and to prevent the crystal from sliding forward the whole apparatus is tilted slightly, the tendency to slide being thus opposed by gravity.

A glass plate is used as a reflecting surface and is mounted upon a brass frame with levelling-screws and springs. This brass frame is carried on the end of a screw of pitch 1 mm., the other end of which carries a circular scale by means of which the traverse of the reflector is measured. With this arrangement it is possible to ensure that the reflecting surface shall be perpendicular to the screw and thus remain parallel to the vibrating surface of the crystal as the screw is turned.

Furnace. The experimental tube telescopes into a second iron tube of greater length round which several turns of ¼-in. copper tubing are wound. Over this is wrapped an enveloping sheet of tinplate, providing a core upon which the wire of an electric furnace is wound. The copper tubing is connected at one end to the gas supply and at the other to the tap at the fixed end of the experimental tube. While observations were in progress a slow flow of gas was maintained through the coil, the gas becoming heated in the process and being then passed through the interferometer tube. The temperature of the gas in the latter was read at the same time. This arrangement maintained a small pressure-gradient from the experimental tube outwards, so preventing a possible leakage of air inwards.

Gas-supply. This was obtained from cylinders as supplied commercially; carbon dioxide was also obtained from a Kipp's apparatus. The only impurity deliberately removed was water vapour, which was disposed of by passing the gases through sulphuric acid and over calcium chloride. Before readings were taken the gas was allowed to flow fairly quickly for about half an hour through the apparatus, to drive out air, and then the flow was cut down until a bubble passed every second or so through the sulphuric acid.

Measurements of wave-length were made in carbon dioxide at frequencies of 41·5, 94, 98, 111 and 145 kc./sec. Richards and Reid's results for 92 kc./sec., published while this work was in progress, are included in the diagram for the sake of comparison. The curve drawn in figure 2 through their points is not that which they regard as the best. Results for nitrous oxide and sulphur dioxide have been confined to two frequencies, viz., 94 and 111 kc./sec. All velocities have been reduced to 0° C. by the usual reduction formula assuming Charles's law. This reduction enables changes in velocity other than that due to change of density of the gas with temperature to be studied.

There are three factors depending on the form of the apparatus which may affect the measured wave-lengths and should be considered. There is first the finite width of the tube, the effect of which is only appreciable however when the tube is narrow and the frequency low. In the present case the Kirchhoff correction from this cause amounts to only one part in 10,000. It has therefore been neglected.

We may take the effects of temperature on the pitch of the screw and on the frequency of the quartz together. The wave-lengths are measured in terms of the pitch of the brass screw at the bush in the fixed end of the experimental tube. As the temperature rises the pitch of the screw increases, making the distance between successive nodes in the stationary wave system appear too small. The temperature coefficient of frequency of the quartz is negative and therefore tends to oppose the other change. The coefficient of reduction of frequency with rise of temperature as measured by Gibbs and Thatte$^{(6)}$, being of the same order as that of the expansion of brass, the two effects have been held to be self-compensating.

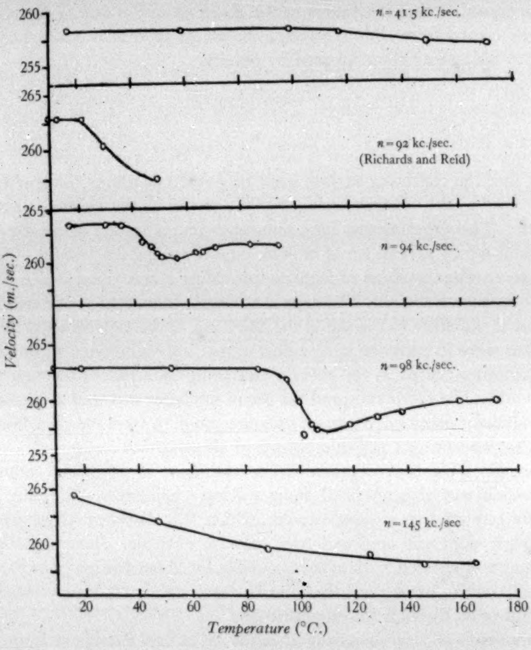


Figure 2. Reduced velocities in carbon dioxide at various frequencies (n).

The results for carbon dioxide are given in table 1 and figure 2; and those for nitrous oxide and sulphur dioxide in table 2.

Table 1. Velocity in carbon dioxide.

θ (°C.)	λ (mm.)	V (m./sec.)	V_0 (m./sec.)
Frequency 41·5 kc./sec.			
17	6·40	266·5	258·5
60	6·86	285·5	258·5
100	7·25	302·0	258·5
118	7·42	309·0	258·0
149	7·65	318·5	257·0
172	7·87	327·5	257·0
190	8·09	336·5	258·5
Frequency 92 kc./sec.*			
10	—	268·5	263·0
14	—	270·5	263·0
22	—	273·5	263·0
30	—	274·5	260·5
50	—	280·0	257·5
Frequency 93·8 kc./sec.			
17	2·895	271·5	263·5
31	2·990	280·0	263·5
37	3·000	281·0	263·5
44	3·010	282·0	262·0
48	3·020	283·0	261·5
50	3·030	284·0	261·0
51	3·030	284·0	260·5
56	3·045	285·5	260·5
64	3·090	289·5	261·0
67	3·105	291·0	261·0
71	3·125	293·0	261·5
84	3·195	299·5	261·5
95	3·240	303·5	261·5
Frequency 98 kc./sec.			
10	2·740	268·5	263·0
20	2·785	272·5	263·0
54	2·940	288·0	263·0
86	3·075	301·0	263·0
97	3·115	305·0	262·0
101	3·110	304·0	260·0
103	3·080	302·0	257·0
106	3·105	304·0	258·0
108	3·110	304·5	257·5
120	3·160	309·5	258·0
130	3·205	314·0	258·5
139	3·245	318·0	209·0
160	3·340	327·0	260·0
174	3·400	333·0	260·0
188	3·450	338·0	260·0
Frequency 111 kc./sec.			
17	2·445	272·5	264·0
37	2·545	282·5	265·0
47½	2·590	287·5	265·5
60	2·655	294·5	267·0
72	2·70	300·5	267·5
Frequency 145 kc./sec.			
15	1·870	272·0	264·5
49	1·955	284·0	262·0
89	2·055	298·5	259·5
127	2·155	313·0	259·0
147	2·200	319·5	258·0
166	2·250	327·0	258·0

* Results due to Richards and Reid(2).

Table 2. Velocity in sulphur dioxide and nitrous oxide.

n (kc./sec.)	θ (°C.)	λ (mm.)	V (m./sec.)	V_0 (m./sec.)
93·8 (SO_2)	18	2·325	217·5	211·0
	25	2·360	221·5	212·0
	35	2·405	225·5	212·0
	47	2·455	230·0	213·0
	57	2·495	234·0	213·0
	67	2·550	239·0	214·0
	76	2·590	243·0	214·5
	88	2·630	246·8	214·0
	93	2·645	248·0	214·0
	108	2·685	251·5	212·5
111 (SO_2)	20	1·995	221·5	213·5
	37	2·045	227·0	213·0
	52	2·110	234·5	215·0
	54	2·120	235·5	215·0
	79	2·195	243·5	216·5
	89	2·215	246·0	213·5
	100	2·240	248·5	212·5
	111	2·270	252·0	212·5
	124	2·230	259·5	215·5
	133	2·280	264·0	216·5
93·8 (N_2O)	19	2·910	273·0	263·5
	35	2·990	280·0	263·5
	45	3·035	284·5	263·5
	56	3·075	288·5	263·0
	66	3·135	294·0	263·5
	77	3·180	298·0	263·0
	95	3·250	304·5	262·5
	109	3·305	310·0	262·0
	128	3·380	316·5	261·5

§ 3. DISCUSSION OF RESULTS

It is proposed to confine discussion to the results for carbon dioxide, since the other two gases do not show so marked a dispersion within the small range of frequency covered. The chief interest in the curves centres round the relatively sharp falls of velocity at certain temperatures for the sources having frequencies near 10^5 c./sec. It is interesting to note that a similarly-shaped curve can be obtained from Smith and Hitchcock's⁽⁷⁾ results for the variation of the dielectric constant ϵ of ice with temperature. By plotting $\epsilon^{-\frac{1}{2}}$, which is a quantity corresponding to a velocity, against temperature for a frequency of 300 c./sec. a curve is obtained not unlike those for supersonic frequencies of 94 and 98 kc./sec. Possibly the theoretical interpretation of the two results will be found eventually to move along parallel lines.

The temperature at which the sharp fall in velocity takes place obviously increases with the frequency nearly in linear fashion, although the number of curves available hardly warrants an attempt to connect these two variables by means of an empirical formula. By trying parameters of the type $n - n_0/(\theta - \theta_0)$, where θ is the temperature and n the frequency, θ_0 and n_0 being constants, it is possible to bring the maxima of dispersion into coincidence, but the effect on the remaining portions of the curves is not so fruitful as the corresponding process in the pressure results⁽³⁾. The curve for 145 kc./sec. shows a gradual fall to a value close to that for 41·5, i.e. towards that for audio-frequencies. This form of curve satisfies the theoretical deductions of Richards and Reid and it may be that their conclusions are only valid for the higher frequencies in the supersonic gamut.

§ 4. ACKNOWLEDGMENTS

The author wishes to tender his sincere thanks to Dr E. G. Richardson for his stimulating interest in this work and advice while it was in progress; to Prof. W. E. Curtis, F.R.S., for putting every facility of the laboratory at his disposal; and to the Education Committee of Durham County Council, whose award of a Senior Exhibition rendered the work possible.

REFERENCES

(1) Kneser. *Ann. Phys.*, Lpz., **11**, 777 (1931).
(2) Richards and Reid. *J. chem. Phys.* **2**, 198 (1934).
(3) Railston and Richardson. Page 533 of this volume.
(4) Sheratt and Griffiths. *Proc. roy. Soc.* **147**, 292 (1934).
(5) Pierce. *Proc. Amer. Acad. Arts Sci.* **60**, 271 (1925).
(6) Gibbs and Thatte. *Phil. Mag.* **14**, 682 (1932).
(7) Smith and Hitchcock. *J. Amer. chem. Soc.* **54**, 4621 (1932).

532.612.4

THE SURFACE TENSION OF A MOVING WATER SHEET

By W. N. BOND, M.A., D.Sc., F.Inst.P., Lecturer in Physics
in the University of Reading

Received December 22, 1934. *Read April* 5, 1935.

ABSTRACT. The direct impact of two cylindrical jets of liquid causes the liquid to emerge radially in a plane at right angles to the common axis of the jets. The maximum diameter of the disc of moving liquid so formed is shown to depend on the surface tension of the liquid.

A method of measuring surface tension is developed and is applied to measure the surface tension of water, giving

$$73{\cdot}8_3 \pm 0{\cdot}1_3 \text{ dyne/cm. at } 15^\circ \text{ C.}$$

The liquid surface is renewed about 80 times per second, contamination being thus prevented. The only other method in which such rapid renewal of the surface takes place is the Rayleigh oscillating-jet method. The angle of contact is not involved in Rayleigh's method, nor in the present method.

§ 1. INTRODUCTION

The method of measuring surface tension that is described in this paper has two advantages. In the first place, the experiment does not depend on the angle of contact. Secondly, the liquid surface on which the experiment is performed is completely renewed about 80 times a second, and is consequently prevented from becoming contaminated. Of the methods that have previously been proposed, the only one in which rapid renewal of the surface takes place is the vibrating-jet method that was originally used by the third Baron Rayleigh(1) and later by Pedersen(2) and by Bohr(3). That method also is independent of the angle of contact.

Before considering the details of the present method, it will be well to describe its essential features. Let us suppose that liquid can be squirted from a nozzle of suitable shape, so that instead of emerging as a cylindrical jet it emerges radially in all directions in a plane. In this way a flat disc of liquid will be obtained, as indicated in figure 1. If the rate of supply of liquid at the centre is kept constant, we expect from continuity that the liquid disc will get thinner and thinner as we recede from the centre. However, if the disc is thin near to the centre, it may be regarded as almost parallel-sided at points that are not near to the centre. If the liquid were all at rest (and if its weight could be neglected) it would require forces acting outwards at its periphery to maintain it in equilibrium. Such forces are not applied externally, but are caused by the loss of outward momentum of the moving

Q, ρ liquid. If the volume of liquid supplied in unit time be Q, if its density be ρ and its
v initial velocity v, then the momentum of the liquid entering the wedge of inde-
α finitely small angle α, figure 1, in unit time is $(\alpha/2\pi)\,Q\rho v$, measured radially outwards. If the liquid came to rest at the outer edge of the disc, it would exert an outward force of $(\alpha/2\pi)\,Q\rho v$. The outward force required to keep the sheet in
R, S equilibrium is $2\alpha RS$, where R is the radius of the outer edge and S is the surface tension of the liquid surface. Hence the maximum radial distance to which the

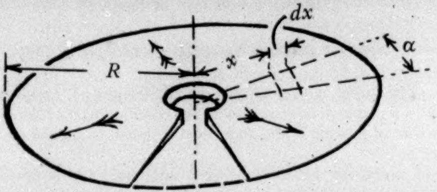


Figure 1.

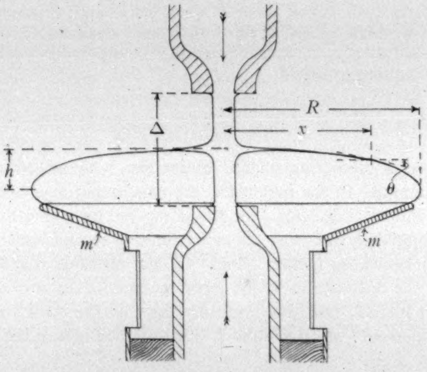


Figure 2.

moving sheet can extend (if no other forces are applied at its outer edge) is given by $2\alpha RS=(\alpha/2\pi)\,Q\rho v$; that is to say, $R=Q\rho v/4\pi S$. Hence, if there were no correction factors to apply, the surface tension would be given by

$$S=\frac{Q\rho v}{4\pi R} \qquad \text{......(1)},$$

where R is the maximum radius of the moving liquid sheet.

In the experiments that I have performed, two equal cylindrical nozzles were arranged facing one another with their common axis vertical, as shown in figure 2. Water was passed through the nozzles, forming a pair of nearly cylindrical jets.

These two jets meet at some point on the common axis, the water then diverging radially in a plane at right angles to the axis of the system. In order that the weight of the disc of liquid shall be small in comparison with the momentum of the liquid delivered by either nozzle in unit time, we must have

$$r^2 \ll S/g\rho \qquad \text{......(2),}$$

where r is the internal radius of the cylindrical nozzle and g is the gravitational acceleration. In the case of water, therefore, the effect of gravity will be small when $r = 0{\cdot}05$ cm., but will be large when $r = 0{\cdot}5$ cm.

r, g

Experiments were carried out with nozzles of radius as small as 0·04 cm., and a sensibly plane disc of water was obtained. The peculiar shape of the edge of the disc is indicated in figure 3. The liquid accumulates in a roll *aa* at the edge of the disc, but irregular cusps *c* rather frequently occur. A spray of drops is shot off tangentially from the two edges of each cusp; and sometimes drops *b* of liquid form on the roll and move slowly along it or drop off below. The cusps may be due to slightly more liquid emerging in some directions in the plane than in others. If a pin or the tip of a knife be inserted into the water sheet a similar cusp is formed.

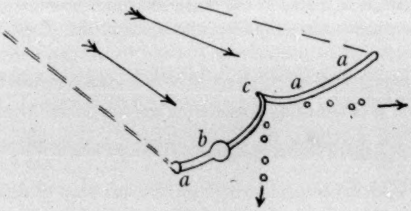


Figure 3.

Measurements of the diameter of the disc showed that, for a particular pair of nozzles, R is proportional to Q^2. The values of Q^2/R for different pairs of nozzles were found to be proportional to r^2. These two results can be deduced from equation (1). Thus, if the velocity in the disc be assumed equal to the mean velocity in the nozzle, we have $Q = 2\,(\pi r^2 v)$, and hence

$$S = \frac{Q^2 \rho}{(2R)\,\pi^2\,(2r)^2} \qquad \text{......(3).}$$

The external radius of the disc was found to be given approximately by equation (3). It was not possible to test the equation very exactly, because the cusps in the outer edge of the disc prevented precise measurement of R.

With nozzles of rather larger radius ($r = 0{\cdot}06$ to $0{\cdot}15$ cm.), the effect of gravity is still small, but not negligible. With the common axis of the nozzles vertical, the horizontal disc curls down under the action of gravity, in the way indicated in figure 2.* The maximum diameter of the bubble-shaped sheet of moving water was susceptible of much more precise measurement than was the maximum diameter of the flat disc, with cusped edges. Calculations and measurements were therefore carried out for such cases. The object of the investigation being to develop a method for the measurement of surface tension, it was not necessary to calculate exactly the effect of gravity on the water sheet, provided that this effect was kept fairly small.

---

* This diagram is drawn for $r^2 = 0{\cdot}2\,(S/g\rho)$. In the case of water, that implies a nozzle whose internal radius $r = 0{\cdot}12$ cm.

§ 2. THEORY OF THE METHOD

x, v

θ

Let us consider a part of the sheet, figure 1, of indefinitely small angle α. At a point at a distance x from the axis let the velocity be v, in a direction making an angle of θ with the horizontal, figure 2. From continuity we find that the thickness of the sheet is $Q/2\pi xv$. Considering a portion of the wedge lying between x and $x+dx$, and equating the rate of gain of downward momentum to the nett effect of the surface tension forces acting at the inner and outer boundaries of the element, together with the weight of the liquid in the element, we obtain

$$\frac{\alpha}{2\pi} Q\rho . d\,(v \sin\theta) = 2\alpha S . d\,(x \sin\theta) + \frac{\alpha}{2\pi} . \frac{Q\rho g}{v} . \frac{dx}{\cos\theta} \qquad \ldots\ldots(4).$$

A similar equation is obtained by considering the horizontal momentum:

$$\frac{\alpha}{2\pi} Q\rho . d\,(v \cos\theta) = 2\alpha S . d\,(x \cos\theta) - 2\alpha S \frac{dx}{\cos\theta} \qquad \ldots\ldots(5).$$

Provided that r^2 is small compared with $S/g\rho$, the last term of equation (4) is small. We may therefore approximate by assuming v to be independent of the value of x *in this small term*. Substituting the value of $dx/\cos\theta$ from equation (5) in equation (4) and integrating we obtain

$$Q\rho\,[v \sin\theta] = 4\pi S\,[x \sin\theta] + \frac{Q\rho g}{v} \left\{ [x \cos\theta] - \frac{Q\rho}{4\pi S}\,[v \cos\theta] \right\}.$$

u_1

u_2

Let the velocity near the central horizontal part of the sheet (*not in the jet*) be u_1, and let the velocity at $x=R$, where $\theta=90°$, be u_2. The rate of gain of downward momentum for the whole of the upper part of the sheet is then equal to either side of the equation

$$Q\rho u_2 = 4\pi SR + \frac{Q^2\rho^2 g u_1}{4\pi S v} \qquad \ldots\ldots(6).$$

h

The effective value of v in the correction term for the weight of the sheet must be between u_1 and u_2 in value. The buoyancy due to the air is quite negligible; and we shall assume that the frictional and inertia effects due to the air are also so small as to be of no importance. In that case u_2 can be deduced in terms of u_1 and the difference in level of the top and outer edge of the sheet, which we will denote h, figure 2. Thus $u_2 = \sqrt{(u_1{}^2 + 2gh)}$.

u

The velocity u_1 will be slightly greater than the velocity u, in the jet, because the pressure in the jet is greater than atmospheric on account of the surface tension of the cylindrical surface. This excess pressure is equal to S/r. Hence, by Bernoulli's theorem we have $\frac{1}{2}\rho u_1{}^2 = \frac{1}{2}\rho u^2 + S/r$. Collecting these results we obtain

$$u_2 = \sqrt{(u^2 + 2gh + 2S/\rho r)}.$$

Before substitution of this result in equation (6), there is one further correction to apply to that equation. It has so far been assumed that no vertical force is applied to the sheet where the water enters it from the two jets, but this is not quite true. If the centre of the liquid disc is stationary at some point on the common axis of the jets, the pressure just below the central point of zero liquid velocity must be equal to the pressure just above that point. By Bernoulli's theorem we find that this is equivalent to saying that, if the jets had not been on the same axis, the velocities in the two jets would have been equal and opposite at points at the same level. The velocity at the lower nozzle will therefore be slightly greater than at the upper, by an amount corresponding to the difference in level. Consequently the mass of liquid emerging from the lower jet per second will be slightly greater than that emerging from the upper jet. When the two jets impinge, they meet with equal and opposite velocities, but there is a nett rate of supply of upward momentum.

If the difference in level of the tips of the two nozzles be small in comparison with the head of liquid that is effective in giving kinetic energy to the liquid in the jets, the nett rate of supply of upward momentum to the centre of the disc of liquid may be written $\pi r^2 \rho g \Delta$, where Δ is the difference in level between the two nozzles, figure 2. Equation (6) may now be modified and becomes

Δ

$$\pi r^2 \rho g \Delta + Q\rho u \sqrt{\left(1 + \frac{2gh}{u^2} + \frac{2S}{\rho r u^2}\right)} = 4\pi SR + \frac{Q^2 \rho^2 g u_1}{4\pi S v} \qquad \text{......(7).}$$

In view of the fact that $Q = 2(\pi r^2 u)$, equation (7) may be simplified, most second-order corrections being neglected. We obtain

$$S = \frac{Q^2 \rho}{(2R - r)\, \pi^2 (2r)^2} - \frac{r^2 g \rho}{2}\left(\frac{S'}{S} - \frac{\Delta}{2R} - \frac{h}{R}\right) \qquad \text{......(8).}$$

Here S' represents the approximate value of S that is given by equation (3) in which there are no correction terms.

S'

It will be useful to recall the origin of the various correction factors that have now been applied to the original equation (3). The term $r^2 g \rho S'/2S$ is due to the weight of the liquid disc. The term $(2R - r)$ owes the r to the pressure in the jets being more than atmospheric because of surface tension. The term in Δ is due to the nett upward momentum supplied by the jets. Finally, the term in h may be ascribed to the downward acceleration of the liquid as it passes round the curved edge of the disc of liquid. In the present experiments, the only correction that exceeded 2 per cent was that due to the weight of the film. As this correction amounted to more than 5 per cent in some of the experiments, it was carried to the second order in equation (8). If an accuracy of $\frac{1}{4}$ per cent had not been sought, equation (8) could have been simplified by writing $S' = S$. By an investigation of the solution of equations (4) and (5) it was found that h/R is approximately equal to $r^2 g \rho / S$. When this estimate did not quite agree with the experimental determinations of h, the mean of the two methods of estimation was used in evaluating the correction factor h/R of equation (8).

Before the experimental work is described, two other possible sources of error will be considered. In the first place, there might be an appreciable conversion of kinetic energy into heat as the liquid flows from the jets into the sheet. It would be very difficult to calculate the exact amount of this error, but it can be shown that the fractional error produced in the estimate of S is of the order $\mu/vr\rho$, where μ is the viscosity of the liquid. In the present experiments we have approximately

$$\mu = 0{\cdot}01 \text{ gm.cm.}^{-1} \text{ sec.}^{-1}$$

$$v = 200 \text{ cm. sec.}^{-1}$$

$$r = 0{\cdot}08 \text{ cm.}$$

$$\rho = 1 \text{ gm. cm.}^{-3}$$

Hence the error produced in the estimates of S is probably of the order of 0·06 per cent and can be neglected.

The possible source of error that still remains to be considered is that due to the velocity in either jet not being quite the same at points near the surface and near the centre of the jet. This error will be discussed in the following section of the paper.

§ 3. APPARATUS AND EXPERIMENTS

The general arrangement of the apparatus is shown in figure 4. The two nozzles were placed with their axes vertical, and held in position by four clamps. If the disc of water is horizontal, as at *a*, figure 4, the tangent plane at its centre being at right angles to the two jets, and if it remains at right angles when the disc moves up or down to a new position, then the two nozzles are accurately aligned. This method of testing the alignment is quite satisfactory, though the adjustments that have to be made to achieve accurate alignment are rather troublesome. It would be of considerable advantage to have the clamps fitted with fine adjustments giving movements in two horizontal directions.

Concentric with the lower nozzle, and fastened to it, was a conical sheet of metal *m*, figures 2 and 4. The function of this is to make the water sheet stable. The metal cone gives a rigid boundary to which the lower part of the water sheet can cling by surface tension. The water in the lower half of the sheet no longer meets in a rather jumbled way near the supply pipe to the lower nozzle. Moreover, the lower part of the conical metal sheet is pierced by several large holes, to ensure that the pressure inside the water sheet shall be atmospheric. Care must be taken that these holes shall not become covered by films of moving water. It might be possible to prevent the formation of such films by having the tops of the holes fitted with baffles.

Round the whole of the nozzle system was a metal can, to collect the water and deliver it to a glass vessel so that it might be weighed. The can was open at the top and had a slot down one side for the lower clamps and supply pipe to pass through. After removal of the support from below the can, the latter could be lowered clear of the nozzle system, to enable adjustments to be made.

A mirror scale *n*, engraved in millimetres, was placed horizontally on the top of the can, figure 4. The scale being viewed from above, the maximum diameter of the water sheet could be measured in several directions two or three times during the course of any experiment.

Tap water was used in the present experiments. The water passed through two constant-level tanks arranged in series, before it came to the apparatus. Consequently the water was supplied at a constant rate, and few air bubbles reached the nozzles. This method of supplying the water was satisfactory, but devices such as that described by Pedersen$^{(2)}$ might be better.

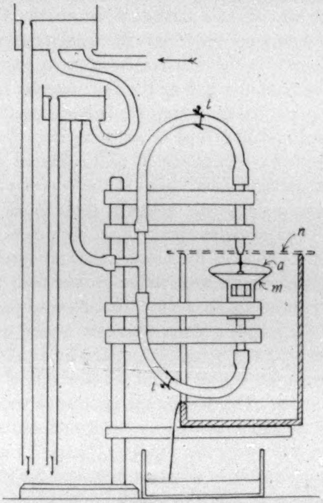


Figure 4.

From the second constant-level tank the water passed to a **T** piece. On the rubber tubes that connected the **T** piece to the two nozzles were screw clips *t*, figure 4. By a slight adjustment of these clips the discharge through the two nozzles could be controlled so that the water sheet floated freely about half way between the two nozzles. If this condition were not fulfilled, the nett upward momentum supplied to the sheet by the pair of jets would not be known, and equation (8) would not be applicable.

It is essential that each nozzle shall deliver a jet which is nearly cylindrical, the velocity of the liquid being nearly the same throughout the jet. It is known that these conditions can most nearly be realized by having a short nozzle consisting of a well-rounded entrance ending in an extremely short cylinder. The method finally adopted for making the nozzles was as follows. A piece of thick-walled capillary glass tubing was selected, the bore of which was nearly circular and of the desired diameter. One end of the tube was closed in a blow-pipe flame and blown out into a small bulb. Then the end of the bulb was blown out, and sealed onto the end of a wider glass tube. Finally the capillary was cut off at a point very near to where it began expanding to form the bulb, so that the required nozzle was left at the end of the wide tube. The cut end may require a little grinding.

A nozzle of rather poorer design was arranged to deliver a jet of water horizontally. The velocity of the water as it left the nozzle was calculated from the curved form assumed by the jet under the action of gravity and also by dividing the rate of discharge by the area of cross-section of the nozzle. The two sets of results did not differ by more than $\frac{1}{4}$ per cent; thus the velocity of the water as it left the nozzle was nearly uniform over the transverse section of the jet.

The internal diameters of the tips of the four nozzles that were used in the final experiments were measured by means of a Hilger travelling microscope. Each nozzle was then arranged to deliver a jet of water horizontally, and the diameter of the jet was measured with the microscope at points distant 10 to 15 jet-diameters from the nozzle. These diameters and those of the corresponding nozzles did not differ by more than about 0·1 per cent, which is about equal to the probable error of the measurements. Tests made at about 50 jet-diameters from the nozzle (but arranged to be at the same level as the nozzle) showed agreement with the other measurements; the probable errors were rather greater than 0·1 per cent, owing to slight sideways movements of the jet during the course of measurement.

Close to the nozzle the jet has a larger diameter, owing to the action of surface tension, the main stream of water being surrounded by a ring-shaped eddy system. In the present experiments we require to find the total rate of supply of momentum from the jet to the liquid disc. This will be the same at all transverse sections of the jet apart from gravitational action and a very small surface tension effect. The rate of supply of momentum may be written $\frac{1}{2}Q\rho u$ or $Q^2\rho/4\pi r^2$, but r must be measured where the jet has become sensibly of constant diameter, for it is only there that the velocity is sensibly the same all over the transverse section.

During the preliminary experiments certain tests were made to confirm the correction terms in equation (8). The term $(2R-r)$ was estimated experimentally as $\{2R-(1{\cdot}0_7 \pm 0{\cdot}1_8)\,r\}$. An approximate test of the Δ term was obtained by finding the rate of discharge of the two nozzles separately. The experimental values of h/R were sometimes greater than and sometimes less than the value predicted from an investigation of equations (4) and (5), namely $r^2g\rho/S$. The mean result from twenty-five experiments was $h/R=(1{\cdot}08 \pm 0{\cdot}06)\,r^2g\rho/S$.

I will now give a short description of how the final experiments were performed. The water was set running through the constant-level tanks, and the nozzles were aligned till the water disc remained horizontal not only when midway between the nozzles but even when it moved up or down to positions nearer one nozzle. The tips of the nozzles were separated by a distance of from $3\frac{1}{2}$ to $5\frac{1}{2}$ jet-diameters, in order that the Δ correction should not be too great, and yet a free floating motion of the water disc should be permitted over a small distance midway between the nozzles. The water-supply to the nozzles was controlled separately till the disc floated freely in this way, the adjustment being slightly altered during the course of the test, if that became necessary. The metal can was raised till it surrounded the whole nozzle system, and was fixed in position. The mirror scale was placed on the top of the can, and a vessel was placed to collect the water as it flowed from the can. The time during which water was collected was measured with a stop watch such that 10 seconds corresponded to one revolution of the hand. A single test lasted for about 8 minutes. During the test, the diameter of the water sheet was measured in 4 or 5 directions, a total of about 12 determinations being made. Directly the water had been collected, its temperature was measured. Before the water was weighed the values of h and Δ were found with a mirror scale. The weight of water was corrected for buoyancy.

§ 4. EXPERIMENTAL RESULTS

Two pairs of nozzles were used, and nine experiments in all were carried out. The results are given in the table. The density of water at the various temperatures was assumed. The estimates of the surface tension were reduced to 15° C. by assuming(4) a temperature coefficient of $-0{\cdot}157$ dyne cm.$^{-1}$ (°C.)$^{-1}$ between 14° and 20° C. Tap water was used in all the experiments.

Table

Nozzle		$2r$ (cm.)	Mass of water (gm.)	Duration of test (sec.)	$2R$, mean (cm.)	θ (°C.)	Δ (cm.)	h (cm.)	h/R mean	$\frac{r^2 g\rho}{S}$
top	bottom									
1*b*	1*a*	$0{\cdot}1399 \pm 0{\cdot}0001_3$	4616	573·3	$4{\cdot}54_3$	17·0	0·58	0·09	0·074	0·064
1*a*	1*b*		4100	505·6	$4{\cdot}61_1$	17·7	0·60	0·18		
1*b*	1*a*		5219	648·5	$4{\cdot}52_9$	19·2	0·56	0·14		
1*a*	1*b*		5425	656·2	$4{\cdot}74_5$	15·4	0·78	0·27		
2*a*	2*b*	$0{\cdot}1851 \pm 0{\cdot}0001_8$	5116	447·9	$5{\cdot}11_8$	17·9	1·00	0·32	0·133	0·116
2*a*	2*b*		5679	506·6	$4{\cdot}95_4$	15·4	0·79	0·30		
2*b*	2*a*		5535	512·8	$4{\cdot}56_5$	15·2	0·77	0·35		
2*a*	2*b*		5468	422·8	$6{\cdot}50_4$	14·4	0·60	0·43		
2*a*	2*b*		5242	411·0	$6{\cdot}27_6$	14·3	0·60	0·43		

$\frac{Q^2\rho}{(2R-r)\,\pi^2\,(2r)^2}$	$-\frac{r^2 g\rho}{2}$	Extra correction on account of $S'/S \neq 1$	$\frac{h}{R}\cdot\frac{r^2 g\rho}{2}$	$\frac{\Delta}{2R}\cdot\frac{r^2 g\rho}{2}$	Temperature correction	S_{15}	
75·16	−2·40	−0·06	0·17	0·31	0·31	73·49	73·73 ± 0·10
75·09				0·31	0·42	73·53	
75·30				0·30	0·65	73·96	
75·78				0·39	0·06	73·94	
76·88	−4·18	−0·16	0·52	0·81	0·45	74·32	73·96 ± 0·13
76·52				0·67	0·06	73·43	
77·09				0·71	0·03	74·01	
77·21				0·38	−0·09	73·68	
77·87				0·40	−0·11	74·34	

The second half of the table gives the values of the correction terms of equation (8). Each of the nine experiments has a probable error of about $\pm\frac{1}{3}$ per cent, due almost entirely to uncertainty in the value of R. The probable errors given in the last column do not include the uncertainty due to the measurement of $2r$. When amended to take this into account, and also to take account of the small discrepancy between the two methods of estimating h/R, the values become

$$\left.\begin{array}{l}\text{For nozzles 1 } a \text{ and } b,\ 73{\cdot}73 \pm 0{\cdot}17\\ \text{For nozzles 2 } a \text{ and } b,\ 73{\cdot}96 \pm 0{\cdot}20\end{array}\right\}$$

giving a final estimate of

$$S\,(15^\circ\ \text{C.}) = 73{\cdot}83 \pm 0{\cdot}13 \text{ dyne/cm.}$$

This value may be compared with the values collected from many sources by Pedersen, by Bohr and by Warren. In particular we have Pedersen's own determination

74·30

that of Bohr (reduced to 15° C.)

72·78

and the value that Warren takes as the best mean of thirteen determinations by various workers

73·65.

The present estimate compares quite satisfactorily with these values. I will only make two comments. Firstly, it is unlikely that the surface tension of tap water, such as was used in the present experiments, differs appreciably from that of distilled water, for no difference was detected by Pedersen or by Bohr. Secondly, I consider the present method worthy of comparison with the Rayleigh oscillating-jet method, even if the present work happens to contain some defects, so far undiscovered, in the details of the theory or experiments.

§ 5. ACKNOWLEDGMENTS

In conclusion I would like to thank Prof. J. A. Crowther, in whose laboratories the work was carried out, and Dr Paul White for their most helpful suggestions. My thanks are also due to Mr J. Burgess, the laboratory steward, for help in making apparatus.

REFERENCES

(1) Rayleigh. *Proc. roy. Soc.* **29**, 71 (1879).
(2) Pedersen, P. O. *Philos. Trans.* A, **207**, 341 (1907).
(3) Bohr, N. *Philos. Trans.* A, **209**, 281 (1909).
(4) Warren, E. L. *Phil. Mag.* **4**, 358 (1927).

538.214:546.655.3

MAGNETIC SUSCEPTIBILITY OF CERIUM CHLORIDE IN AQUEOUS SOLUTION AND ITS VARIATION WITH TEMPERATURE*

By JEANNE LIQUIER-MILWARD, Docteur ès Sciences, Ph.D.

Received January 30, 1935. *Read in title April* 5, 1935.

ABSTRACT. Measurements on the paramagnetic susceptibility of cerium chloride $CeCl_3.6H_2O$ in aqueous solutions have been made over a wide range of concentrations (8 to 37·5 per cent of the anhydrous salt) and for the temperature-interval 7° to 70° C. by means of a slightly modified Quincke ascension method. The variation of the molecular susceptibility with temperature is found to follow, for the interval considered, a Curie-Weiss law

$$\chi_c (T+\Delta)=C, \text{ with } C \text{ equal to } 0{\cdot}762.$$

The value of C appears to be unaffected by the variation of the concentration of the solution. On the other hand, Δ varies with the number of ions present in the unit of volume and increases from 45 to 63 for the range of concentration considered.

The effective magneton number μ_e or $14{\cdot}07 \sqrt{(\chi_c T)}$ of the ion Ce^{+++} at room temperature is calculated and is found to vary, for this same interval of concentrations, from 11·41 to 11·16. If, on the other hand, the magneton number is deduced from the modified form of the above formula, $\mu_w = 14{\cdot}07 \sqrt{\{\chi_c (T+\Delta)\}}$, a value is obtained which is independent of the concentration and of the temperature and is equal to 12·36.

These results are compared with the values obtained experimentally by other workers and with the theoretical numbers deduced by Hund and Van Vleck from the study of the spectroscopic ground states of the rare-earths group.

Some measurements on the densities of the solutions under investigation are included.

§ 1. INTRODUCTION

Since the recent development of the theory of paramagnetism, the study of the magnetic behaviour of the ions of the rare-earths group appears to be of special interest. This is due to the fact that in all these ions the incomplete group of electrons responsible for the paramagnetism, instead of being in an outer shell as is the case for the ions of the first transition series, lies deeper in the ion (4_4 electrons), which implies that the magnetic carriers are less liable to be affected by the field from neighbouring ions and molecules. The rare-earths ions are thus the best substitute available for the ideal monatomic gas for which the theoretical formulae have been established. Consequently, one might expect the theory to apply to them and a good agreement to be found in their case between experimental data and theoretical calculations.

---

* Part of a thesis submitted for the degree of Ph.D. to the Birmingham University on October 1, 1934.

χ

The new quantum mechanics leads$^{(1)}$, for the susceptibility χ of an ideal monatomic gas, to the well-known formula

$$\chi=\frac{N\mu^2}{3kT}+N\alpha \qquad \ldots\ldots(1),$$

N, μ
T, k, α

where N is the number of molecules per cm.^{3}, μ the permanent magnetic moment of the molecule, T the absolute temperature, k the Boltzmann constant and $N\alpha$ the joint contribution to the expression for χ of the high-frequency elements of the paramagnetic moment and of the diamagnetic effect.

This general formula takes different forms according to the value of the width of the multiplet intervals as compared to kT. When the multiplet intervals are large in comparison with kT, as is assumed to be the case for the ions of the rare earths, formula (1) for the susceptibility becomes

$$\chi=\frac{Ng^2\beta^2 J(J+1)}{3kT}+N\alpha \qquad \ldots\ldots(2),$$

J, β
g

where J is the resultant of the orbital and of the spin moments, β the Bohr magneton and g the Lande splitting factor. For most of the ions of the rare earths, moreover, the second term $N\alpha$ is small and negligible when compared with the temperature-dependent term $Ng^2\beta^2J(J+1)/3kT$.

Hund$^{(2, 3)}$, assuming the electron-distribution proposed by Stoner, predicted the values of the respective quantum numbers J, S, L, and the value of g for all the ions of the rare-earths group; he computed the corresponding Weiss magneton numbers from the formula $\mu_w=4{\cdot}97Jg\sqrt{\{(J+1)/J\}}$ and compared the theoretical values thus obtained with the experimental values deduced from the measurements of magnetic susceptibilities then available. The agreement appears at first sight to be very satisfactory with two or three exceptions. Hund's values for samarium and europium are much too low. Van Vleck's estimations, however, calculated with the assumption that the multiplet intervals for these two ions are not very large compared with kT, are in better agreement with experiment.

However, the experimental values of the magneton numbers recorded by Hund were all calculated on the assumption that the susceptibilities of all rare-earths ions varied with the temperature according to a Curie law, as gadolinium was then known to do; they correspond to the effective Weiss magneton numbers as defined by the formula

μ_e

$$\mu_e=4{\cdot}97\sqrt{(3\chi_e kT/L\beta^2)}$$

β, χ_e, L

where β is the Bohr magneton, χ_e the susceptibility per gram-ion, and L the Avogadro number. In numerical terms

$$\mu_e=14{\cdot}07\sqrt{(\chi_e T)}.$$

The measurements of susceptibilities have been made only at room temperature, in most cases in the crystal state.

In reality, accurate measurements of the susceptibilities at different temperatures are absolutely necessary if a really convincing comparison is to be made between experimental data and the theoretical values deduced from formula. Some such measurements have been made recently by Cabrera and Duperier(4) for some of these ions, most of them on oxides and on the anhydrous sulphates in the solid state. Most of the ions under investigation are found to follow the Weiss generalization of the Curie law

$$\chi(T+\Delta)=C.$$

Recently, too, measurements have been made for samarium in $Sm_2(SO_4)_3.8H_2O$ by S. Freed(5), and for neodymium by P. W. Selwood(6).

In Hund's tables, cerium (Ce^{+++}) appears as one of the ions for which there is a definite discrepancy between the theoretical and the experimental values of the magneton number, and an investigation of the influence of temperature on the paramagnetism of this ion seemed to us to be of interest. It will be remembered that Ce^{+++} is the simplest amongst the paramagnetic ions of the group and has only one electron in the incomplete group, the structure of its different shells of electrons being assumed to be the following:

Ce^{+++} electronic shell	1_1	2_1	2_2	3_1	3_2	3_3	4_1	4_2	4_3	4_4	5_1	5_2
Number of electrons	2	2	6	2	6	10	2	6	10	1	2	6

This configuration is the one which serves as basis to Hund's deduction of the moment.

In the present paper, and as a preliminary to a more extensive investigation on cerium salts, a study is made of the paramagnetic susceptibility of one salt of cerium, $CeCl_3.6H_2O$, in aqueous solution, for different concentrations and for different temperatures so as to ascertain whether the variation of the susceptibility of the trivalent ion Ce^{+++} follows a Curie law $\chi T=C$, or whether there is a molecular field, which would lead to the Weiss formula $\chi(T+\Delta)=C$.

§ 2. PREVIOUS MEASUREMENTS ON THE SUSCEPTIBILITY OF CERIUM SALTS

Some estimates of the magnetic susceptibility in a few compounds of trivalent and quadrivalent cerium are already available. The older measurements are those of St. Meyer, Du Bois and Du Bois and Liebknecht. St. Meyer(7) investigated, amongst other salts of cerium, cerium chloride in aqueous solution, for a concentration of $n=4{\cdot}4$ moles per litre and found for the gram susceptibility of the solution the value $5{\cdot}20 . 10^{-6}$. In a more recent paper(8) the magneton number of Ce^{++++} is given as equal to 0·8 but there is no mention of any experiment on Ce^{+++}; the value given for the magneton number of Ce^{+++} is in reality that for Pr^{++++}, assuming the Sommerfeld-Kossel rule that ions with equal numbers of electrons have the same susceptibilities.

Du Bois(9) found for the molecular susceptibility of cerium sulphate the value $2330 . 10^{-6}$. Du Bois and Liebknecht(10) give for the molecular susceptibility χ_m of $CeCl_3$ and $CeBr_3$ respectively $\chi_m=2430 . 10^{-6}$ and $\chi_m=2400 . 10^{-6}$. In their experiments the values taken for the susceptibility of water is rather different from the one which is now assumed, and the sample of cerium investigated was not free from lanthanum.

χ_m

More recently Decker$^{(11)}$ measured the susceptibility of cerium nitrate and found that it was the same for the solid state as for the salt in solution, the susceptibility per gram-ion being $1890 . 10^{-6}$, which yields a value 10·46 for the magneton number μ_e. Decker's measurements were made only for small values of the concentration and the correction applied to the molecular susceptibility in order to obtain the susceptibility of the ion Ce^{+++} does not include the underlying diamagnetism of the ion itself; the magnetic number calculated from these measurements is much too low.

So far no attempt had been made at measuring the susceptibilities at different temperatures.

I. Zernike and C. James$^{(12)}$ investigated, amongst other sulphates, $Ce_2\,(SO_4)_3 . 5H_2O$ in the solid state, and as they had not the adequate equipment they tried to obtain a qualitative insight into the temperature coefficient by varying the temperature of the room in which the experiments were made. This was only a rough attempt which did not lead to any definite conclusions. These authors give $\chi_m = 2377 . 10^{-6}$ for the susceptibility per gram-atom at 20° C.

E. H. Williams$^{(13)}$ found a remarkably low value for cerium probably because he had, in the sample investigated, a mixture of Ce_2O_3 and CeO_2, where Ce is tetravalent and diamagnetic. As a whole the main difficulty in these measurements is to find a very pure chemical.

In his first investigation on the rare earths, Cabrera$^{(14)}$ found the value 11·89 for the magneton number of Ce^{+++} from measurements on the solid hydrated sulphate $Ce_2\,(SO_4)_3 . 8H_2O$, the experiments being made only at room-temperature. As there was good agreement between the results obtained for the solid and for the dissolved states, that is to say for very different degrees of magnetic dilution, most of the authors cited above concluded that even in the crystal the ions are magnetically free and that the 4_4 electrons are not influenced by any molecular field. In an addition to their earlier work Cabrera and Duperier$^{(4)}$ have recently measured different oxides and anhydrous sulphates of the rare-earths ions for a range of temperature; no value, however, is given for cerium.

The different values given for the magneton number of Ce^{+++} in Weiss magnetons are summarized in table 1.

Table 1. Weiss magneton number of Ce^{+++}

No. of 4_4 electrons	Basic level	Jg	Theoretical value of μ_e		Experimental value of μ_e			
			Hund	Van Vleck	Cabrera	St. Meyer	Zernike and James	Decker
1	${}^2F_{\frac{5}{2}}$	$\frac{5}{2} . \frac{6}{7} = 2{\cdot}14$	12·5	12·7	11·8	13·8	11·7	10·4

On the whole, the values given for the magnetic susceptibility of cerium do not agree and no measurement has been made of its temperature coefficient. Also, in works cited above, the corrections which have been applied for the diamagnetism of the acid radical are not stated and no correction is mentioned for the underlying diamagnetism of the ion Ce^{+++} itself.

§ 3. METHOD OF EXPERIMENT

The Quincke ascension method, as improved and described by Piccard(15), has been adopted, the temperature of the solution under investigation being controlled by circulating water from a thermostat. A minor improvement, however, has been introduced in the present apparatus with regard to the water jacket surrounding the capillary tube at the level of the meniscus.

As is well known, a U-tube containing the liquid investigated is employed in the Quincke method, one limb being of wide and the other of narrow bore. The narrow tube passes between the poles of an electromagnet, the meniscus of the liquid being placed in a region where the field is uniform. When the current is applied, a displacement h of the meniscus is observed. If H is the value of the field and if the magnetic susceptibility of the air is neglected, the gram susceptibility of the liquid is given by the relation: h, H

$$\chi = 2hg/H^2.$$

Actually, the displacement of the liquid is not observed but the meniscus is brought back to its original position by raising the reservoir R, figure 1, by means of a suitable micrometric device. The various readings are thus all made strictly for the same value of the field. The value of H is not measured directly, the apparatus being calibrated instead by measurements on distilled water, the value $-0{\cdot}72 . 10^{-6}$ being taken as its specific susceptibility χ_w with respect to air at 20° C. and 76 cm. If, for the same value of the field H, h_w is the displacement observed for water, χ is given by the formula: χ_w h_w

$$\chi = \frac{h}{h_w}\chi_w.$$

The U-tube employed is similar to the one used by Piccard but has a better jacket arrangement, see figure 1. In Piccard's apparatus the space available between the pole pieces being small (5 mm.), the water jacket is interrupted at the level of the meniscus so as to avoid having too capillary a tube; as a result the region of the liquid which is under investigation is the one which is unprotected. In the present experiments we used an interspace of 8 mm. which made possible the use of a continuous water jacket J round the meniscus portion of the tube, the bore of the inside tube being still sufficiently wide. The meniscus is observed through the layer of circulating water, and this can easily be done as long as the circulating water is free of air bubbles.

The temperature of the water is recorded in (*a*) before being admitted into the thermostatic jacket and in (*b*) immediately after leaving it by means of two precise thermometers which are graduated in tenths of a degree C. and are fitted in glass water jackets. As there is necessarily a certain length of rubber tube between (*a*) and (*b*), the water cools down slightly and there is a small difference between the readings of the two thermometers. For temperatures not exceeding 40° this difference does not exceed 0·4°, but it may attain 0·8° for higher temperatures. The average value between the two readings is taken as the value of the temperature T at the level of the meniscus and can be considered to be known to 0·1° or at the worst 0·2°.

The electromagnet employed is of the Weiss type and gives a field of 15500 G. when a current of 5 A. is applied. The meniscus of the liquid being always brought back to the same position, the uniformity of the field is not fundamental; however, an exploration of the field with a Gauss fluxmeter and a search coil proved it to be uniform over a considerable portion of the space between the pole pieces. In order to avoid the error which can be introduced by the hysteresis of the iron, a second reading is taken with the field reversed and the average of the two results is taken
h' as the apparent value h' of the displacement of the liquid in the field. The setting of the micrometric screw which was used to adjust the level of R could be repeated to 0·02 mm. for all the measurements recorded below.

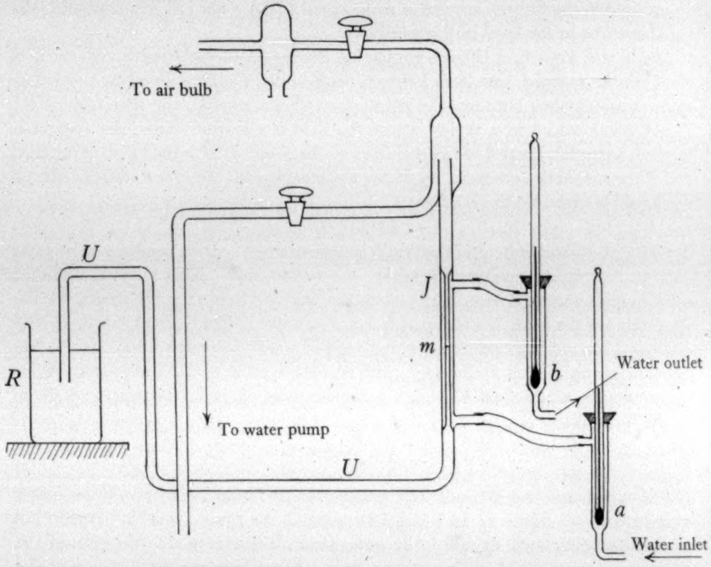


Figure 1. Detail of U-tube and water jacket arrangement.

In order to obtain the real value of h which figures in the above formula, some corrections are necessary. The outside diameter of the tube U being an appreciable fraction of the diameter of the reservoir R a correction must be applied for the part of the tube which is immersed when the level is adjusted. Moreover, the formula

applies only when the temperature of the liquid is uniform. In the present apparatus there is an appreciable difference between the temperature T in m and that T
in R which is kept at room-temperature t; consequently, the difference of density t
in m and R must be taken into account. In the present case, the respective diameters of the tube and of the reservoir being 6 mm. and 48 mm., the true rise of the liquid is given by the following expression:

$$h=h'.\frac{65}{64}.\frac{d_t}{d_T}.$$

In order to check the method of experiment and the working of the apparatus, a first set of measurements has been made on a solution of anhydrous manganese sulphate in water, for which the variation of susceptibility with temperature has already been accurately measured by Theodorides⁽¹⁶⁾. This author, using the Gouy method, has found that anhydrous $MnSO_4$ follows a Curie-Weiss law $C=\chi(T+\Delta)$, with a value of the Curie constant $C=4{\cdot}267$, which yields for the magneton number
of manganese $\mu_w=29{\cdot}05$. The chemical used in the present measurements has been μ_w
analyzed so as to ascertain that it was free from water. The concentration of the solution investigated was 18·05 per cent. The value of the correction for the diamagnetism of the anion has been taken as equal to 37.10^{-6} since $\chi_{SO_4}=-37.10^{-6}$. The value found for the Curie constant C is 4·136 and the corresponding magneton number μ_w is $14{\cdot}07\sqrt{4{\cdot}136}$ or 29·02. These two numbers are in remarkably good agreement with the results of Theodorides cited above.

§4. MATERIALS

The cerous chloride, $CeCl_3.6H_2O$, used in the present investigation was obtained from two sources. A first sample A was provided by The British Drugs Supply, a second B was obtained from Johnson Matthey & Co. Chemical analysis did not show any traces of lanthanum, praseodymium and neodymium which are likely to be present with cerium. The observation of the absorption spectrum has in both cases been made on a layer of the saturated solution of chloride 1 cm. thick. For sample A some lines belonging to praseodymium and neodymium showed very faintly; no absorption spectrum could be seen for B.

Careful analysis of the crystals showed the presence of $6H_2O$ per molecule, as indicated by the formula for sample A. Sample B contained a small quantity of extra water which was evaluated to within 1 per cent. To prepare each solution of a given concentration, the necessary weight of salt is added to the corresponding weight of distilled water; the value of the concentration which can then be calculated is always checked by chemical analysis; the Cl is determined by precipitation of silver chloride and the Ce by precipitation of the oxalate. The concentration of
each solution in anhydrous salt is given as c in grams per gram of solution. c

§5. EXPERIMENTAL RESULTS

When the gram susceptibility χ of the solution is known, the gram susceptibility of the anhydrous salt is computed from the following Wiedemann formula, assuming that the susceptibility of the solution is the sum of the susceptibilities of its components:

$$\chi_s = \frac{\chi + (1 - c)\, 0{\cdot}72 \,.\, 10^{-6}}{c}.$$

The values obtained for the anhydrous salt are then converted to molecular susceptibilities by multiplying by the molecular weight 246·6. Corrections must then be applied for the diamagnetism of the acid radical and for the underlying diamagnetism of the paramagnetic ion itself. For the acid radical, the value $\chi_{Cl} = -20 \,.\, 10^{-6}$ given by Pascal (as deduced from molecular susceptibility) was taken. The value of the diamagnetism of the ion itself is more difficult to compute. Pauling gives for the molecular diamagnetism of Ce^{++++}, $\chi_{at} = -33 \,.\, 10^{-6}$. Since the diamagnetic susceptibility is given by the Langevin formula

$$\chi_{at} = -(e^2/4mc^2) \sum_n \overline{r_1^2},$$

$\overline{r_1^2}$

n

where $\overline{r_1^2}$ is the time mean square of the radius of the resolved electronic orbit perpendicular to the field and n the number of electrons, it can be assumed that the diamagnetism of Ce^{+++} is slightly larger, for this ion has an extra outer electron. We adopted as a reasonable estimate of the susceptibility the value $-35 \,.\, 10^{-6}$ which is also in good agreement with the Thomas and Fermi numerical formula

$$\chi_{at} = -10^{-5} \,.\, Z^{\frac{1}{3}}$$

Z

in which Z is the atomic number.

Thus, the total quantity which is to be added to the molecular susceptibility to obtain the susceptibility per gram ion is $+95 \,.\, 10^{-6}$.

χ_s

χ_m, χ_c

In the following tables, χ_s is the specific or gram susceptibility of the anhydrous salt, χ_m the molecular susceptibility and χ_c the susceptibility per gram-ion corrected for the diamagnetism of the crystal water and the diamagnetism of both anion and cation. The values of χ_c are believed to be known within 1 per cent.

In figure 2, the reciprocal $1/\chi_c$ of the ionic susceptibility is plotted against the absolute temperature, and in each case a straight line is obtained. The experimental points do not lie off the straight line by more than 1 per cent. Assuming that the Curie-Weiss law is followed, we have for each experimental point the relation

$$\frac{1}{\chi_c} = \frac{T}{C} + \frac{\Delta}{C}.$$

The values of C and Δ corresponding to each solution have been calculated by application of the principle of least squares to the four sets of relations thus obtained. For each of the four solutions investigated, values of Δ are found which are not zero. Consequently it may be concluded that the variation of susceptibility of Ce^{+++} with temperature does not follow a Curie law as it has so far been assumed to do, but that there is a molecular field whose constant remains at about 45 even for the most dilute solution investigated ($c = 0{\cdot}0857$).

A notable variation of Δ with the concentration is observed, Δ increasing from 45 for $c = 0{\cdot}0857$ to 63 for $c = 0{\cdot}375$, as is shown in figure 2. Consequently, the suscepti-

Table 2

$CeCl_3.6H_2O$; $c=0·0857$; $d_{283°}=1·083$; $\Delta=+45$.

T (°K.)	d	$\chi_s.10^6$	$\chi_m.10^6$	$\chi_c.10^6$	$\frac{1}{\chi_c}$	$\chi_c.(T+\Delta)$
280·6	1·084	9·06	2234	2329	429	0·758
287·0	1·082	8·89	2192	2287	437	0·759
287·2	1·082	8·89	2192	2287	437	0·759
308·5	1·073	8·34	2057	2152	465	0·761
330·8	1·062	7·77	1916	2011	497	0·756
342·0	1·057	7·55	1862	1957	511	0·757
343·4	1·056	7·58	1869	1964	509	0·763

Table 3

$CeCl_3.6H_2O$; $c=0·1240$; $d_{283°}=1·123$; $\Delta=+46·4$.

T (°K.)	d	$\chi_s.10^6$	$\chi_m.10^6$	$\chi_c.10^6$	$\frac{1}{\chi_c}$	$\chi_c.(T+\Delta)$
282·1	1·123	9·08	2239	2334	428	0·767
283·6	1·122	9·03	2227	2322	430	0·766
291·6	1·119	8·71	2148	2243	445	0·758
299·1	1·116	8·51	2098	2193	455	0·758
307·0	1·114	8·31	2049	2144	466	0·758
315·0	1·110	8·23	2029	2124	472	0·768
323·5	1·106	8·00	1972	2067	484	0·765
328·6	1·103	7·83	1931	2026	493	0·760
331·5	1·101	7·80	1923	2018	495	0·762
334·4	1·099	7·74	1909	2004	499	0·763
336·9	1·097	7·70	1898	1993	502	0·763

Table 4

$CeCl_3.6H_2O$; $c=0·2247$; $d_{293°}=1·234$; $\Delta=+48·7$.

T (°K.)	d	$\chi_s.10^6$	$\chi_m.10^6$	$\chi_c.10^6$	$\frac{1}{\chi_c}$	$\chi_c.(T+\Delta)$
281·7	1·239	8·86	2185	2280	439	0·753
288·1	1·2365	8·78	2165	2260	442	0·761
288·5	1·2365	8·78	2165	2260	442	0·762
289·3	1·236	8·71	2148	2243	446	0·758
289·4	1·236	8·70	2145	2240	446	0·757
290·8	1·235	8·67	2138	2233	448	0·758
293·1	1·234	8·64	2131	2226	449	0·761
294·8	1·233	8·58	2116	2211	452	0·759
298·8	1·232	8·43	2079	2174	460	0·756
302·0	1·231	8·40	2071	2166	462	0·760
316·9	1·223	7·97	1965	2060	485	0·753
319·7	1·221	7·96	1963	2058	486	0·758
324·2	1·218	7·85	1936	2031	492	0·757
331·5	1·213	7·71	1901	1996	501	0·759
332·0	1·213	7·69	1896	1991	502	0·758
336·6	1·210	7·52	1853	1948	513	0·751
341·3	1·207	7·47	1842	1937	516	0·755
343·1	1·205	7·48	1845	1940	515	0·760
346·4	1·203	7·32	1805	1900	526	0·751
348·0	1·202	7·41	1827	1922	520	0·762
350·4	1·200	7·37	1817	1912	523	0·763

Table 5

$CeCl_3.6H_2O$; $c=0·3750$; $d_{284°}=1·453$; $\Delta=+63$.

T (°K.)	d	$\chi_s.10^6$	$\chi_m.10^6$	$\chi_c.10^6$	$\frac{1}{\chi_c}$	$\chi_c.(T+\Delta)$
283·7	1·453	8·62	2095	2190	450	0·759
292·5	1·449	8·42	2076	2171	461	0·771
307·6	1·441	8·01	1976	2071	483	0·768
316·8	1·437	7·88	1943	2038	491	0·774
322·6	1·433	7·71	1902	1997	501	0·770
325·1	1·432	7·70	1898	1993	502	0·773
339·3	1·424	7·38	1819	1914	522	0·770

Table 6

c	C	Δ	μ_w or 14·07 $\sqrt{\{(T+\Delta).\chi_c\}}$
0·0857	0·759	+45·0	12·26
0·1240	0·762	+46·4	12·28
0·2247	0·757	+48·7	12·24
0·3750	0·772	+63·0	12·36

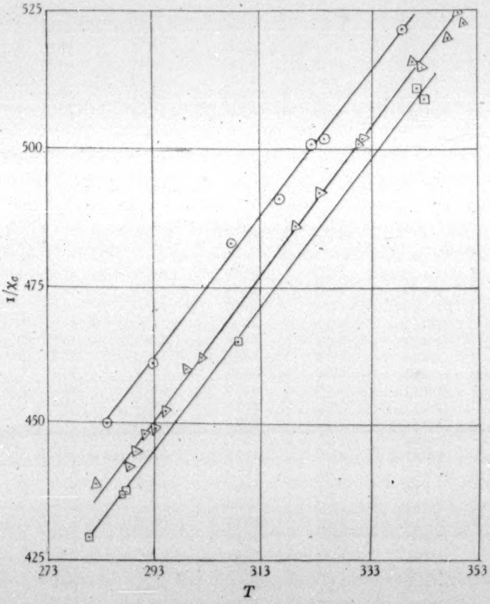


Figure 2. Variation of $1/\chi_c$ with temperature and concentration. $p=0·0857$ ⊡; $p=0·2247$ ◬; $p=0·3750$ ⊙.

bility and the corresponding effective magneton number are a function of the concentration. The fact that, for the high magnetic dilution corresponding to $c = 0{\cdot}0857$, Δ has the value 45 shows that the molecules of the solvent itself exert an appreciable influence on the magnetic carriers.

For the four solutions investigated, C has been found to be constant within 1·5 per cent, its average value being 0·762.

It must be noted that a similar change in the Curie point without a corresponding change of the Curie constant has been observed also on neodymium compounds by Selwood(6) who found that Δ, as in the present case, is a function of the concentration.

If we calculate the effective magneton number at room-temperature as defined by the relation $\mu_e = 14{\cdot}07\sqrt{(\chi_c T)}$, the following values are found:

c	Moles per litre	μ_e at 19° C.
0·124	0·56	11·41
0·2247	1·12	11·27
0·3750	2·21	11·16

The value corresponding to $c = 0{\cdot}1240$ is in good agreement with the ionic magneton number, $\mu_e = 11{\cdot}8$, given by Cabrera from measurements at room-temperature on the solid sulphate. As a whole the numbers obtained are much smaller than the theoretical numbers given by Hund and by Van Vleck (respectively 12·5 and 12·7). It is noteworthy that if the magneton number is deduced from experiment by using the modified form of the above formula

$$\mu_w = 14{\cdot}07\sqrt{\{\chi_c(T+\Delta)\}},$$

a value is obtained for μ_w which is independent of the concentration and the temperature and is equal to $14{\cdot}07\sqrt{0{\cdot}772}$ or 12·36. This last result is in good agreement with Hund's theoretical value 12·5; however, one must bear in mind that the theoretical value has been computed with the assumption that there is no molecular field, and consequently this is only a spurious agreement.

As the range of temperature available for aqueous solutions is very limited, further experiments are being made on solutions of $CeCl_3$ in other solvents. Some measurements are being made also to evaluate the molecular field constant in solid compounds of cerium.

§ 6. ACKNOWLEDGMENT

In conclusion I wish to express my thanks to Prof. S. W. J. Smith for granting me full research facilities and for his interest throughout the work.

REFERENCES

(1) Van Vleck J. H. *Electric and Magnetic Susceptibilities*, p. 227 (1932). Oxford.
(2) Hund, F. *Linienspektren*, Chap. v.
(3) —— *Z. Phys.* **33**, 855 (1925).
(4) Cabrera, B. and Duperier, A. *C.R. Acad. Sci.*, Paris, **188**, 1640 (1929).
(5) Freed, S. *J. Amer. chem. Soc.* **52**, 2702 (1930).
(6) Selwood, P. W. *J. Amer. chem. Soc.* **55**, 3161 (1933).
(7) Stefan Meyer. *Ann. Phys.*, Lpz., **69**, 236 (1899).
(8) —— *Phys. Z.* **26**, 51, 478 (1925).
(9) Du Bois, H. *Berichte*, 3344 (1899).
(10) Du Bois, H. and Liebknecht, O. *Ann. Phys.*, Lpz., **4**, 189 (1900).
(11) Decker, H. *Ann. Phys.*, Lpz., **79**, 324 (1926).
(12) Zernike, I. and James, C. *J. Amer. chem. Soc.* **48**, 2827 (1926).
(13) Williams, E. H. *Phys. Rev.* **29**, 218 (1927).
(14) Cabrera, B. *C.R. Acad. Sci.*, Paris, **180**, 668 (1925).
(15) Piccard, A. *Arch. Sci. phys. nat.* 118^e année, 4^e période, **35**.
(16) Theodorides, Ph. *C.R. Acad. Sci.*, Paris, **171**, 715 (1920).

546.303:539.15:530.145

A DISCUSSION OF THE TRANSITION METALS ON THE BASIS OF QUANTUM MECHANICS

By N. F. MOTT, M.A.

Received March 15, 1935. *Read May* 3, 1935.

ABSTRACT. The magnetic susceptibilities and electrical conductivities of cobalt, nickel and palladium and of their alloys with copper, silver and gold are discussed on the basis of the quantum theory of metals. It is shown that the number of electrons in the outermost *s* state must be about 0·6 per atom in the transition metals and one in the noble metals; certain magnetic properties of the alloys are explained on the basis of this fact. A quantum-mechanical explanation of the relatively high resistance of the transition metals is given and is shown to be supported by measurements of the resistance of alloys.

§ 1. INTRODUCTION

Wigner and Seitz⁽¹⁾ have recently shown that the quantum theory is capable of accounting quantitatively for the cohesive forces in a typical metal (sodium); the purpose of this paper is to give a qualitative discussion, based on the same theory, of some properties of the transition metals and in particular of the elements nickel and palladium, which come immediately before copper and silver in the periodic table. After an introductory discussion of the electronic structure of these metals and a comparison between the binding forces of nickel and copper, we shall consider (i) the saturation moments of the ferromagnetic metals and alloys, and the reason why they are not equal to integral numbers of Bohr magnetons per atom (§§ 2 and 3); (ii) the paramagnetism of palladium and of its alloys with copper, silver and gold (§§ 4 and 5); and (iii) the electrical conductivity of the transition elements, and the reason for their low conductivity as compared with that of the noble metals. The resistance of alloys of the transition metals with copper, silver and gold is also discussed (§ 6).

To obtain an approximate solution of the Schrödinger equation for the electrons in a metal we may start from one or other of two models; we may picture the electrons as bound to individual atoms (method of Heitler and London and of Heisenberg) or we may think of them as belonging to the crystal as a whole (method of Bloch). It is to be emphasized that these two models do not correspond to different physical states of the crystal; both lead to wave functions of the whole system which are approximations to the true wave function. Insulators, as well as conductors can be treated, to a certain degree of approximation, by the use of either model. In this paper we shall use that of Bloch not only for the outermost *s* electrons, which are responsible for the cohesion and for the conductivity, but also for the *d* electrons in the incomplete shells, which are responsible for the ferromagnetism of nickel and for the high paramagnetism of palladium. We do not, of course, consider that the Bloch approximation is the better of the two when the overlap between the wave functions of two atoms is as small as it is for, say, the 3*d* states of nickel(2); we use it because with this model one may include, in the zero order of approximation, the possibility that an atom may be ionized, or may contain a non-integral number of electrons.* To take these possibilities into account in the Heisenberg model would be much more complicated.

$\psi_{\mathbf{k}}$

In the Bloch approximation, each electron is described by a wave function $\psi_{\mathbf{k}}$, which is a solution of the Schrödinger equation

$$\frac{\hbar^2}{2m}\nabla^2\psi_{\mathbf{k}}+(E_{\mathbf{k}}-V)\,\psi_{\mathbf{k}}=0$$

V
$\mathbf{k}$
$E_{\mathbf{k}}$
N_A
n

where V is the potential of a periodic field extending throughout the crystal. The subscript $\mathbf{k}$ denotes the state of the electron; owing to the exclusion principle, not more than two electrons may be in each state. The allowed energies $E_{\mathbf{k}}$ lie in zones or bands; if the atoms are a long way apart, the zones are extremely narrow, but as the atoms are brought nearer together the zones broaden out and may overlap. Each zone corresponds to a stationary state of a single electron in the field of the isolated atom;† the number of states in any zone is always equal to the statistical weight of the atomic state to which it corresponds, multiplied by the number of atoms N_A in the crystal. For the transition elements we are interested in the bands which correspond to the *ns* and $(n-1)$ *d* states of the free atom, n being the principal quantum number of the outermost *s* electron, i.e. 4 for nickel and 5 for palladium. The *ns* band contains one state per atom and the $(n-1)$ *d* band five. Two electrons may be in each state. Since these elements have 10 electrons which must be shared between the two bands, and since the ferromagnetism shows that in nickel at any rate the 3*d* band is not full, it is clear that the two bands must overlap.

In the elements nickel, palladium and platinum the states with the electron configurations‡ $\{(n-1)\,d\}^{10}$ and $\{(n-1)\,d\}^9\,(ns)^1$ have energy differing by an amount considerably less than the binding energies of the crystals (approximately 4 electron-volts per atom). The energies are given in table 1 in electron volts.

Table 1

	Nickel	Palladium	Platinum
Energy required to raise atom from the lowest state with the configuration $\{(n-1)\,d\}^9\,(ns)^1$ to that with $\{(n-1)\,d\}^{10}$	1·4	−0·81	0·76

---

* Cf., for example, the work of H. Jones on alloys which obey the Hume-Rothery electronic rule, reference (3).

† The simplified model with which we are working does not include exchange forces, and therefore from our point of view the singlet and triplet states of given quantum number in, for instance, a two-electron system must be regarded as the same state.

‡ The $\{(n-1)\,d\}^8\,(ns)^2$ configuration, which is the lowest for a free nickel atom, is of small importance in the metal, as will appear below.

Thus the work required to take an electron from the *s* state to the *d* state is small and, in so far as it is legitimate to refer to the energy of a single electron, we may say that the energy of an electron in the $(n-1)$ *d* state is nearly the same as in the *ns* state.

For our discussion of the transition metals the essential assumption is that the interaction energy between the *d* shells of neighbouring atoms is small. As has been pointed out by Slater⁽²⁾ this is probable, because the overlap between the wave functions is small. The *d* band will therefore be narrow (less than 1 e-V.) and its mean energy will not be displaced very much from the position of the *d* state in the free atom. The *s* electrons will therefore be responsible for nearly all the cohesion. The number of electrons in the *s* band will not change with temperature, except by a small quantity of the order of magnitude kT/(binding energy per atom).

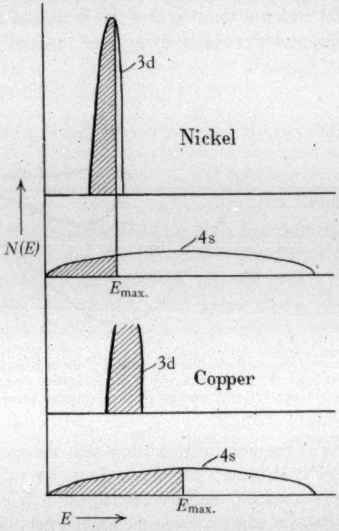


Figure 1. Density of electronic states $N(E)$ as a function of the energy E for a transition metal (nickel) and noble metal (copper). The shaded areas represent occupied states. The total number of states, $N(E)\,dE$, is $5N_A$ in a *d* band and N_A in an *s* band, where N_A is the total number of atoms.

We denote by⁽⁴⁾ $N(E)\,dE$ the number of states in a given band with energy between E and $E+dE$. Figure 1 shows the general form of $N(E)$ plotted against E for two such metals as nickel and copper which, having nearly equal atomic volumes so that they are adjacent in the periodic table, probably have similar internal fields. There are five times as many states in the *d* band as in the *s* band. In copper the N, E

p

3d band is full* with 10 electrons and the 4s band half full with one electron per atom. In nickel, with one electron fewer, there will be a certain number p of electrons per atom in the s band and an equal number p of positive holes in the d band.† If one assumes that the energies of the bands in copper and nickel are the same, it is clear from the figure that p is less than unity, i.e. that nickel will have *fewer s* electrons than copper. If copper is alloyed with nickel we should on the basis of this model expect that the number of s electrons will increase only very slightly with increasing copper content until the d band is full, when the ratio of the number of copper to that of nickel atoms will be $p : (1-p)$. In the next sections, we shall refer to the abundant experimental evidence showing that this is so, and that analogous results hold for palladium alloyed with noble metals.

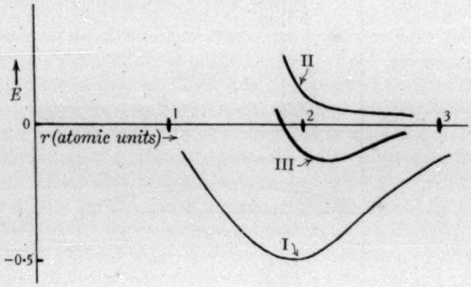


Figure 2. Energy E of the electrons in copper, as a function of the atomic radius r, calculated by the method of Wigner and Seitz. The unit of energy is the ionization potential of hydrogen; (I) the energy of the lowest s state; (II) the energy of the occupied state with maximum energy; (III) the mean energy of the electrons.

The considerations of the next section show that for nickel the number p of electrons in the s band is about 0·6 per atom. In order to gain a better understanding of the energies involved, we have carried out a Wigner-Seitz calculation for the neighbouring element copper, using the atomic field calculated by Hartree. We have only made the calculation to the degree of approximation adopted in Wigner and Seitz's first paper, and have therefore made no allowance for the correlation between the positions of electrons either with parallel or with antiparallel spin; and we have made no allowance for the effect on the 3d electrons of the alteration in charge-density of the 4s electron. The Fermi energy was calculated by a method similar to that of Slater(6) and found to be about 1·4 times greater than the value given by the Sommerfeld formula.

The results of this calculation are shown in figure 2. The curve (I) gives the energy of the electron in the lowest state, (II) that of the electron in the highest

---

[*] The considerations set forth in this paper show that if this were not the case copper would have the low conductivity and large magnetic susceptibility characteristic of a transition metal.

[†] Further evidence of the existence of unoccupied d states with large values of $N(E)$ is afforded by the work of Veldkamp(5) on the fine structure of L-edges in the X-ray absorption spectra of Ta and W.

state, and (III) the mean energy—i.e. the energy of the lattice. The abscissae r are defined by

$$\tfrac{4}{3}\pi r^3 = \text{the atomic volume.}$$

The calculation gives $r = 3{\cdot}15$ atomic units, as against an observed value $r = 2{\cdot}66$, and a binding energy 0·14 Rydberg units, as against 0·25 observed. The agreement with experiment is not good, but the calculations illustrate our point that the electrons with maximum energy have *greater* energy than in the free atom, so if the $3d$ shell had in it any vacant places of energy about the same as in the free atom (shown by the horizontal line), the electrons would be in a state of lower energy if they occupied those places.

If these ideas are correct, we should expect the binding energies of nickel, palladium and platinum to be greater than those of copper, silver and gold. The values in table 2 are given by Grimm and Wolff$^{(7)}$.

Table 2. Binding energies (kilocal./gm.-atom)

Nickel	101	Copper	76
Platinum	122	Gold	83

The fact that the work function of all metals is less than the ionization potential of the free atom shows that the top electrons of the Fermi distribution are contributing a negative amount to the cohesion, and will go into the $3d$ shell if there is any room there. From the considerations given above, we should expect the difference between the work function and the ionization potential to be *less* for the transition metals than for the noble metals. The values observed are given in table 3.

Table 3

	Work function (e-V.)	Ionization potentials (e-V.)
Nickel	5·01	7·2*
Copper	4·38	7·68
Palladium	4·96	7·5*
Silver	4·6	7·5
Platinum	6·3	8·9*
Gold	4·8	9·2

§ 2. SATURATION MOMENTS OF THE FERROMAGNETIC ELEMENTS

It is a well known fact that the saturation moments of ferromagnetic elements do not correspond to integral numbers of Bohr magnetons $e\hbar/2mc$ per atom. The saturation moments for the three ferromagnetic elements$^{(8)}$ are as shown in table 4.

It is generally recognized that the ferromagnetism is due to $3d$ electrons. From the point of view of the Bloch theory there will exist in the solid a band of states† corresponding to the 3*d* states of the atom, and another band corresponding to the 4*s*; in nickel the bands must certainly overlap, or the 3*d* shell would be full and the 4*s* band empty, and the metal would be an insulator. There will thus be for each metal a certain number p of electrons per atom in the *s* band, and $10-p$, $9-p$, $8-p$ in the *d* band for nickel, cobalt and iron respectively.

---

* The values given are the energies required to ionize an atom in the lowest state having the configuration $\{(n-1)\,d\}^9\,ns^1$.

† Cf. figure 1.

Bloch⁽⁹⁾, and more recently Wigner⁽¹⁰⁾ have discussed ferromagnetism starting from the model used here; they have shown that under certain conditions the exchange forces may be such that the state of lowest energy is reached when some of the electrons have their spins uncompensated, i.e. when more spins are parallel to a given direction than are antiparallel.

Table 4

Element	Saturation intensity per atom (units of $e\hbar/2mc$)
Iron	2·2
Cobalt	1·7
Nickel	0·6

Table 5

	Number of electrons per atom in 4*s* band
Nickel	0·6
Cobalt	1·7 − 1 = 0·7
Iron	2·2 − 2 = 0·2[*]

We must now enquire why it is that the state of lowest energy is reached when (in nickel) 0·6 electron per atom have their spins uncompensated. There are two possibilities; either a balance is then reached between the exchange forces and the Fermi force,[†] so that the total energy of the 3*d* electrons is then a minimum with respect to the magnetic moment; or all states in the 3*d* shell with given spin-direction are then full. The former assumption has been made by Stoner;[‡] if, however, it is correct, the work required to reduce the intensity of a saturated magnet by an amount δI is proportional to $(\delta I)^2$, instead of to δI, as in the Weiss theory; the first assumption therefore leads at low temperatures to a very rapid decrease of magnetization with increasing temperature, and is thus incompatible with the observations. We therefore assume the second alternative to be true. The work done in decreasing the magnetization by δI is then proportional to δI, as in the Weiss theory, except in so far as electrons may make transitions from the
E_f 4*s* to the 3*d* band; but since the mean kinetic (Fermi) energy E_f of a 4*s* electron is of the order of magnitude of 6 e-V. and increases rapidly with the number of 4*s*

---

[*] For iron this conclusion can hardly be accepted, since 0·2*s* electron is too few to account for the binding energy. Probably the *d* band is split by the structure of the crystal, and one of the sub-bands is saturated.

[†] I.e. the force which keeps the electron spins antiparallel in normal metals.

[‡] See reference (8), p. 431.

electrons (see below), the number which can make such a transition is of the order of magnitude kT/E_f and is therefore negligible.

We thus deduce that the saturation intensity is the maximum possible for the number of positive holes left in the $3d$ band. The number p of electrons in the $4s$ band is therefore as shown in table 5 for the three elements.

We shall see in § 4 that palladium like nickel and cobalt has also about 0·6 electron per atom in the outer s band; the evidence for platinum is less definite.

§ 3. ALLOYS OF NICKEL AND COBALT WITH OTHER METALS

The theory given above accounts at once for the remarkable results of Alder(11), who has found that in the copper-nickel alloys, which have a face-centred cubic structure over the whole range, the saturation moment at low temperatures is decreased by one Bohr magneton for each copper atom which replaces a nickel atom in the lattice. This rule has been verified up to 40 per cent of copper; by extrapolation the saturation intensity will reach zero for an alloy containing 60 per cent of copper, and 40 per cent of nickel. Sadron(12) has found, moreover, that in the alloys of nickel with zinc, aluminium and tin, the replacement of a nickel atom by an atom of one of the three elements mentioned decreases the moment by approximately two, three and four Bohr magnetons respectively.

On the model given above, this is to be explained as follows. In all these alloys the maximum binding energy will be obtained when the number of electrons in the $4s$ band is about 0·6 per atom, as for nickel. The extra electrons will therefore go into the $3d$ band, as long as there is room for them there. But in the $3d$ band, at low temperatures all the states with spin parallel to the direction of magnetization are already occupied (hypothesis of § 2). Therefore the electrons must go into states having the opposite spin. Thus if a nickel atom is replaced by an atom of copper, zinc, aluminium or tin, the saturation intensity of magnetization will be decreased by one, two, three or four Bohr magnetons as the case may be.

This explanation does not necessarily imply that the $3d$ shells of copper, zinc, etc., are to any extent ionized when alloyed with nickel. In zinc, for instance, the $3d$ levels are much lower than in nickel, and so the $3d$ band will split into two; wave functions corresponding to energies in the lower band will be small except in the neighbourhood of zinc atoms. If the number of zinc atoms is $N_A x$ and of nickel atoms $N_A(1-x)$, the number of states in the lower and upper bands will be $10N_A x$, $10N_A(1-x)$ respectively. The lower band will always be full, so that the mean magnetic moment in the neighbourhood of a zinc atom is zero.

We have seen that cobalt and nickel have about the same number of $4s$ electrons. Experiment* shows that the saturation intensity of cobalt-nickel alloys plotted against atomic composition gives a fairly straight line. This shows that the number of $4s$ electrons remains between 0·6 and 0·7 throughout the range, while the magnetic moment of the $3d$ shells remains as great as the number of positive holes allows.

---

[*] Cf. reference (8), p. 532.

Sadron(12) finds that the replacement of one nickel atom by an atom of manganese *increases* the magnetic moment by three Bohr magnetons, this finding being valid up to about 6 per cent of manganese. This is to be expected, because manganese has three less electrons than nickel, and, assuming always that the firmest binding (lowest energy) occurs with 0·6 electron in the 4*s* band, the replacement of a nickel atom by one of manganese will decrease the number of electrons in the 3*d* band by three. If as many electrons in the 3*d* band as possible have their spins parallel to the field, Sadron's result follows.

On the other hand, it is not easy to see why the addition of manganese to cobalt decreases the moment.

The addition of palladium to nickel leaves the saturation moment unchanged, up to about 50 per cent of palladium. This shows that in these alloys the number of electrons in the *s* band is also about 0·6, as for nickel. Similarly, the addition of palladium or platinum to cobalt gives a curve for the saturation moment very like that of cobalt-nickel; the cobalt-palladium alloys have a saturation moment of 0·5 Bohr magnetons per atom *of both kinds* even for only 15 per cent of cobalt(13).

§ 4. PARAMAGNETISM OF PALLADIUM AND PLATINUM

The transition metals palladium and platinum have paramagnetic susceptibilities which vary with temperature much less than a normal paramagnetic substance but are large compared with those of the other metals. This paramagnetism is certainly due to the uncompleted shells: definite evidence to this effect is given below. We consider it extremely improbable that orbital motion is responsible for any significant part of the susceptibility; the gyromagnetic effect shows that in the ferromagnetics, even above the Curie point(14), the magnetism is at any rate mainly due to spin, and thus that the interaction between the spins is sufficient to quench the orbital motion. Now Slater(6) has pointed out that the overlap for the incomplete *d* shells is *less* for the ferromagnetic than for the similar non-ferromagnetic elements. We should therefore expect, a fortiori, that the orbital motion would be quenched for the non-ferromagnetic elements.

We shall therefore assume that the paramagnetism is a spin paramagnetism. Pauli(15) has given a theory of the paramagnetism of free electrons, which may be modified* to apply to the case when the electrons move in a periodic field. The

χ formula for the susceptibility χ is

$$\chi = 2\mu^2 N(E_{\text{max.}}) \qquad \mu = eh/2mc \qquad \ldots\ldots(1),$$

$N(E_{\text{max.}})$ where $N(E_{\text{max.}})$ is the number of electronic states per unit energy range at the surface of the Fermi distribution—i.e. in the occupied state of highest energy. Since $N(E)$ is very large for a narrow band, one can always obtain agreement with experiment by assuming the breadth of the band to be sufficiently small.†

We shall make an estimate of the breadth of the band for palladium, assuming that the *d* shell is full except for 0·6 electron per atom. At the head of a band $N(E)$ will have the form*

---

* Cf. Sommerfeld and Bethe, reference (4), p. 473.

† Cf., for instance, reference (16).

$$N(E)=A(E_0-E)^{\frac{1}{2}} \qquad \text{......(2),}$$

E_0 being the energy of the highest state in the band.

Now the number n of positive holes per unit volume is

$$2\int_{E_{\text{max.}}}^{E_0} N(E)\,dE=\frac{4A}{3}(E_0-E_{\text{max.}})^{\frac{3}{2}} \qquad \text{......(3).}$$

Hence, by equation (1)

$$\frac{n}{\chi}=\frac{\frac{4}{3}(E_0-E_{\text{max.}})}{2\mu^2}.$$

With n equal to 0·6 (atomic volume) this gives for the energy interval between the surface of the Fermi distribution and the top of the *d* band

$$E_0-E_{\text{max.}}=0{\cdot}05\ \text{e-V.} \qquad \text{......(4).}$$

Since we have at present no a priori knowledge of the breadth of the band, it is not impossible that it should be as narrow as this. On the other hand, the specific heat of the free electrons is given, according to the usual theory, by†

$$c_v=\frac{2\pi^2}{3}N(E_{\text{max.}})\,k^2\,T,$$

T

so that, if in palladium and platinum the large paramagnetism were entirely due to a large $N(E)$, there would be a considerable contribution to the specific heat from the *d* electrons. The relation between the specific heat and the susceptibility would be

$$c_v=\frac{\pi^2}{3}\frac{1}{\mu^2}k^2T\chi \qquad \text{......(5),}$$

which gives for palladium ($\chi=64{\cdot}10^{-6}$) and platinum ($\chi=28{\cdot}10^{-6}$) the values 0·009 T and 0·004 T, in calories per gram atom. Such large values are incompatible with the experimental observations, since they give, for palladium at room temperatures, a specific heat greater than R from the electrons.‡

We therefore conclude that, for palladium and platinum, the density of states $N(E)$ for the *d* band, though considerably greater than for free electrons, is not sufficient to give the observed paramagnetism, but that the latter is due to an exchange force of the same type as is responsible for ferromagnetism. We shall therefore assume

$$\chi=2A\mu^2N(E_{\text{max.}}) \qquad \text{......(6),}$$

where A is in general greater than unity, and depends in some unknown way on the distance apart of the atoms.

A

§ 5. PARAMAGNETIC SUSCEPTIBILITY OF ALLOYS OF PALLADIUM WITH COPPER, SILVER AND GOLD AND OF PALLADIUM SATURATED WITH HYDROGEN

Experiments have been carried out by Svensson(17) on the magnetic susceptibility of the copper-palladium and silver-palladium alloys, by Vogt(18) on gold-palladium, and by Aharoni and Simon(19) and by Svensson(20) on palladium saturated with hydrogen. If our interpretation of the paramagnetism is correct, the effect on the susceptibility of substituting an atom of a noble metal is twofold; by altering the interatomic distance the exchange forces, and hence A in equation (6), may be changed; and by filling up the d band $N(E_{\text{max.}})$ will be decreased. Since we are not in a position to calculate A, we shall discuss the change in $N(E_{\text{max.}})$.

---

* Cf. Sommerfeld and Bethe, reference (4), p. 473.

† Cf. reference (4), p. 430.

‡ Note added in proof. Keesom and Clark(32) have recently found for nickel below 4° K. a value for the specific heat of 0·0019 T, which suggests that $E_0-E_{\text{max.}}$ is about 0·2 e-V. for this element.

For any nearly full band of a pure metal, $N(E_{\text{max.}})$ is equal to the cube root* of the number of unoccupied "holes" in the band considered ($4d$ for palladium). If we assume the same to be true for the alloy, this would give us, for an alloy consisting of Nx atoms of copper, silver or gold and $N(1-x)$ of palladium,

$$N(E_{\text{max.}}) = \text{const.} \times (p-x)^{\frac{1}{3}} \qquad \ldots\ldots(7),$$

where p is the number of positive holes per atom in pure palladium. On the other hand, as explained in § 3, if the d levels have different energies in the two atoms concerned, the d band will split into two, and only the upper band, corresponding to the $4d$ states of palladium, will have positive holes. In this case, the number of positive holes per palladium atom is

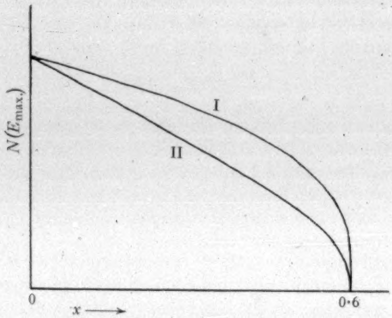


Figure 3. The density of states at the surface of the Fermi distribution (I) from equation (7); (II) from equation (8).

$$\frac{p-x}{1-x},$$

and hence for $N(E_{\text{max.}})$ we should write

$$N(E_{\text{max.}}) = \text{const.} \times (1-x)\left(\frac{p-x}{1-x}\right)^{\frac{1}{3}} \qquad \ldots\ldots(8),$$

$N(E_{\text{max.}})$ according to equations (7), (8) is illustrated in figure 3 with p equal to 0·6.

This formula, moreover, is only accurate if the palladium atoms are arranged in a regular way in the crystal—i.e. if they form a superlattice, which is not in general the case. For a random distribution of the palladium atoms, we should expect the upper limit of the band to be less sharp. In consequence, $N(E_{\mathrm{max.}})$ will not disappear so suddenly with increasing x as equations (7) and (8) suggest.*

---

* This is easily seen from equations (2) and (3).

In the case of hydrogen dissolved in palladium, it is probable that the hydrogen atoms give all their electrons to the $4d$ levels of palladium, since the proton does not replace a palladium ion in the lattice. Thus for Ny hydrogen atoms dissolved in N atoms of palladium, we shall have

$$N(E_{\mathrm{max.}}) = \text{const. } (p-y)^{\frac{1}{3}} \qquad \ldots\ldots(9).$$

Figure 4 shows the observed susceptibility of palladium-silver plotted against x (where $100x$ is the number of silver atoms per 100 atoms of both kinds), and of palladium-hydrogen against y (where $100y$ is the number of hydrogen atoms per 100 palladium atoms). As the equations (8) and (9) lead us to expect, the susceptibility for palladium-hydrogen falls to zero more sharply. Neither curve follows at all exactly the theoretical curve; this may be ascribed to the variation of the exchange force, and to lack of sharpness of the edge of a band mentioned above.

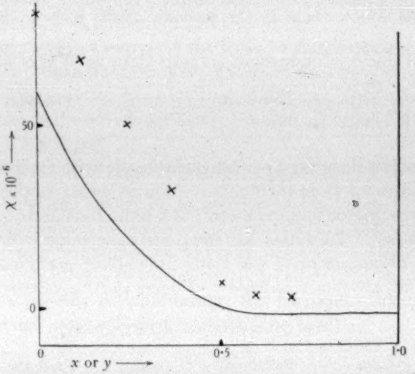


Figure 4. The full lines shows the susceptibility of palladium-silver, and the crosses the susceptibility of palladium + hydrogen. x denotes the ratio of silver to atoms of both kinds; y denotes ratio of hydrogen to palladium atoms.

We deduce that in palladium the number of $5s$ electrons per atom is about 0·5.† That the number (0·5) of s electrons in palladium is less than for nickel (0·6) is to be expected, because in atomic palladium the $(4d)^{10}$ state is the lowest, whereas in nickel the $(3d)^{10}$ state is 1·4 e-V. higher than the lowest state with the configuration $(3d)^9\,(4s)^1$. Therefore in palladium more electrons will go into the d band.

If the view given here of the absorption of hydrogen by palladium is correct, the fact that palladium becomes saturated at the composition Pd_2H is not to be attributed to the formation of a compound, as suggested by Linde and Borelius⁽²²⁾ and others. It must be interpreted as meaning that a hydrogen atom can only enter the lattice if its electron goes into the *d* shells of the surrounding palladium atoms, so that when these *d* shells are full, no more hydrogen can be absorbed. The saturation at a composition near Pd_2H follows from the fact that palladium has about 0·5 positive holes in the *d* shell.

---

* Svensson (17) has found that for the copper-palladium alloys, where the susceptibility disappears at about 50 per cent of copper, the disappearance is sharper for annealed alloys in which the atomic arrangement is ordered than for quenched alloys, where the arrangement, as the electrical conductivity shows, is disordered.

† The relation between the disappearance of paramagnetism of palladium for 50 per cent noble metal or hydrogen, and the disappearance of the ferromagnetism of nickel for about the same amount of copper, was first pointed out in an interesting paper by Dorfmann (21).

§ 6. ELECTRICAL CONDUCTIVITY

All the transition metals are comparatively poor conductors; the conductivities of nickel, palladium and platinum are compared below with the conductivities of the elements that follow them in the periodic table, namely copper, silver and gold.

Table 6. Conductivity σ ($cm.^{-1}\,\Omega.^{-1} \times 10^{-4}$) at 0° C.

Nickel	16·1	Palladium	10·3	Platinum	10·2
Copper	64·5	Silver	66·7	Gold	49·0

According to modern theories, a perfect lattice has infinite conductivity; but when the atoms are displaced from their mean positions owing to thermal motion, the electrons may be scattered and the metal has a finite resistance.

In a recent paper⁽²³⁾ the author has compared the conductivities of the elements for equal mean displacement of the atoms from their positions of equilibrium; Θ the results ($\sigma/M\Theta^2$) are shown in table 7, Θ being the characteristic temperature.

Table 7. $\sigma/M\Theta^2$ (arbitrary units)

Nickel	0·2	Palladium	0·24	Platinum	0·105
Copper	1·05	Silver	1·32	Gold	0·82

In the paper quoted it was suggested that the scattering power of two ions, for given atomic displacement, would only differ by a few per cent for atoms near together in the periodic table, and having nearly equal atomic volume, such as for instance nickel and copper. It was therefore suggested that the difference between the conductivities of two neighbouring metals was due to the different effective* numbers of free electrons. Experimental results on the resistance of dilute solid solutions were quoted to show that this is the case for such metals as silver, cadmium, magnesium, etc., which have no incomplete shells. For nickel and palladium, however, we have found the actual number of electrons in the *s* band to be 0·6 and 0·5, and so the effective number of free electrons should not be less than half that for copper, silver and gold. The reason for the big difference in $\sigma/M\Theta^2$ must therefore be sought elsewhere.

The positive holes in the *d* band make a certain contribution to the conductivity —i.e. they are free to move through the lattice. But since the atomic *d* wave functions do not overlap much, the positive holes will take a much longer time to move from one atom to the next than would be taken by an s electron, and so the contribution to the effective number of free electrons is small. The positive holes, however, will increase the *resistance* in the following way. The resistance of a metal is proportional, among other things, to the number of times per second an electron is scattered—i.e. to the number of times per second that it makes a transition from a state specified by a wave vector $\mathbf{k}$ to any other state $\mathbf{k}'$. Now the probability for such a transition is proportional† to $N(E_{\max.})$, the density of states; for if $N(E_{\max.})$ is big, there are more states into which the electron can jump. In the transition metals, $N(E)$ is big in the d band; *and therefore electrons will jump more frequently from the s to the d band than from one s state to another.* The time of relaxation for such metals is therefore shorter, and the conductivity smaller than for the noble metals, in which only s—s transitions can take place. $\mathbf{k}, \mathbf{k}'$

---

* Cf. Sommerfeld and Bethe, reference (4), p. 378; or Mott⁽²³⁾.

In order to convince ourselves of the truth of this hypothesis, we must show that the transition-probability from a given state in the s band to one in the d band is comparable with that between two states in the d band. The transition-probability between two states with wave functions ψ_k, $\psi_{k'}$ is proportional to the square of a matrix element of the type‡

$$\int \psi_{\mathbf{k}'}^{*}(r) \operatorname{grad} V \psi_{\mathbf{k}}(r)\, d\tau,$$

where V is the potential energy of an electron in the lattice. For $\psi_{\mathbf{k}'}$, the wave function of the final state of the electron, we may take, to a sufficient approximation, the wave function ψ_d^{n-1} in an unperturbed atom in the $(n-1)$ d state. The perturbing energy, grad V, is the change in the potential due to moving an atom through unit distance from its position of equilibrium; the perturbing energy in the neighbourhood of any one atom will therefore be of the form $V, \psi_{\mathbf{k}'}$

$$f(r)\cos\theta.$$

Therefore if $\psi_{\mathbf{k}}$ were an s function (i.e. spherically symmetrical) the transition-probability would be zero. For a wave function at the surface of the Fermi distribution, however, $\psi_{\mathbf{k}}$ will be of the form[(23)] given by

$$\psi_{\mathbf{k}} = A\psi_s^n + B\psi_p^n + C\psi_d^n + \dots \qquad \dots\dots(10),$$

where ψ_s, ψ_p, ψ_d are supposed normalized and have the symmetry of s, p, d wave functions with principal quantum number n and $A \gg B \gg C$. The transition-probability between two states in the s band will be of the order of

$$\left| AB \int \psi_s^n \operatorname{grad} V \psi_p^n \, d\tau \right|^2,$$

and between two states, one in the s band and the other in the d band,

$$\left| B \int \psi_d^{n-1} \operatorname{grad} V \psi_p^n \, d\tau \right|^2.$$

These two expressions contain the same power of the small coefficient B, and may be taken to be of the same order of magnitude.

Evidence of the truth of the hypothesis that the high resistance of the transition metals is due to the large density of states, $N(E)$, in the d band can be derived from the electric resistance of the gold-palladium and silver-palladium series of alloys. Both have the face-centred cubic structure throughout the range of composition, and do not form a superlattice.

---

† Cf. reference (4), p. 519.

‡ Cf. reference (4), p. 512.

R_0
R_t

The resistance of any disordered alloy is made up of two parts; the resistance R_0 due to disorder, and the resistance R_t due to thermal motion of the atoms. The latter may be taken to be given by

$$R_t = T\frac{dR}{dT}.$$

In figure 5 we have multiplied the right-hand side by a constant differing slightly from unity, to obtain the correct resistance of the pure metal.

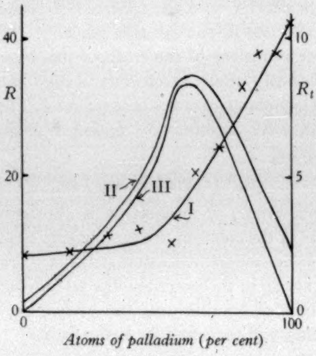


Figure 5. Resistance of palladium-gold and palladium-silver alloys. The crosses denote the experimental values of R_t for palladium-gold; curve (I) is drawn through these points. Curves (II) and (III) show the total resistance of palladium-silver at 20° C. and $-273°$ C. respectively.

Since R_t may be as little as one-tenth of the total resistance, its accurate estimation is difficult. Figure 5 shows the values deduced from measurements of the resistance and temperature coefficient due to Giebel$^{(25)}$, for palladium-gold. The behaviour of palladium-silver is similar.*

According to our hypothesis, if palladium is added to gold, the number p of s electrons will decrease to a limiting value of about 0·55 at 45 per cent of palladium. This should increase the resistance, but since the palladium ion has probably less scattering-power than that of the heavier gold, we may expect the thermal part R_t of the resistance to remain roughly constant up to this composition. For less than 55 per cent of gold, however, the d band has vacant places in it, and we should expect R_t to increase, and the increase in the resistance to be proportional to $N(E_{\text{max.}})$. As figure 5 shows, the variation of $N(E_{\text{max.}})$ that one deduces from this curve is very similar to that given by the paramagnetic susceptibility of palladium-silver, figure 4. Again, the disappearance of $N(E_{\max.})$ at about 50 per cent of silver is not so sharp as the theoretical equation (8) suggests.

---

* Since the total resistance $R_0 + R_t$ is greater for this alloy, the experimental values of R_t are probably less certain.

In figure 5 we show also the total resistance at room temperature of the silver-palladium alloys, and further the resistance at the absolute zero of temperature, obtained by subtracting the observed values of TdR/dT. The curves for gold-palladium and copper-palladium are similar, but the maximum occurs nearer to the 50 per cent composition. For the copper-nickel alloys $^{(26)}$ the dependence on temperature is complicated by the fact that the alloys are ferromagnetic, and the change in the temperature coefficient in the neighbourhood of the Curie point, which for about 50 per cent of copper is in the neighbourhood of room temperature. At low temperatures, however, the measurements of de Haas and Krupkowski$^{(27)}$ show a curve very similar to that of figure 5 with the maximum at 45 per cent of copper. The sharp maximum of the resistance curves is in sharp contrast to the resistance of, say, silver-gold or palladium-platinum at low temperatures, which follow a curve of the type

$$R \propto x(1-x) \qquad \text{......(11)},$$

with a much flatter maximum; cf. figure 6.

We give in table 8 the increase ΔR in the resistance of the noble metals due to admixture of 1 per cent of a transition metal and *vice versa*.*

Table 8. ΔR ($\mu\Omega./\text{cm}^3$)

Gold in palladium	1·0	Palladium in gold	0·407
Silver in palladium	1·28	Palladium in silver	0·436
Copper in palladium	1·27	Palladium in copper	0·82
Copper in nickel	1·1	Nickel in copper	1·25

As the author has pointed out in a previous paper$^{(23)}$, the scattering-power due to 1 per cent of a metal A in solid solution in B is the same as that due to 1 per cent of B in solid solution in A, and for pairs of metals of similar valency and structure (silver-gold or palladium-platinum), the increase of resistance ΔR is the same. If however atoms of a foreign metal can cause transitions from a state in the s band to the d band, we should expect ΔR to be much greater for, say, copper in nickel than for nickel in copper. This is not the case, and, even for the silver and gold alloys with palladium, the difference is only such as may well be accounted for by the fact that silver and gold have *one* s electron per atom in the s band and palladium only about 0·5. It is known from other evidence that the s electron of copper is rather less free than for silver and gold,† so that the effective number of free electrons per atom is rather less than unity. We deduce that the transition probabilities from s to d states in the metal are small, probably smaller than the s—s transition-probabilities. This is to be expected for the following reason: the possibility that in a transition metal an electron will make a transition from the s

---

* The values are due to Svensson$^{(17)}$ and Linde$^{(28)}$, except for copper-nickel, where they are taken from (29).

† From the measurements of the optical constants$^{(27)}$, and also from the fact that $\sigma/M\Theta^2$ is *less* for copper than for silver, in spite of the smaller ion of copper and the larger number of electrons per unit volume.

band to a *d* state under the influence of an atom of copper, silver or gold in solid solution is proportional to

$$|\int \psi_d^{n-1}(r)\,\Delta V\,\psi_{\mathbf{k}}(r)\,d\tau\,|^2 \qquad \ldots\ldots(12),$$

where ψ_d^{n-1} and ψ_k are defined above and ΔV is the change in the potential in the lattice due to the addition of an atom of copper, silver or gold. But this integral will only have a finite magnitude if the wave function ψ_d^{n-1} of the final *d* state of the electron is finite in the region where ΔV is different from zero; i.e. in the immediate neighbourhood of the foreign atom. This will only be so if the *d* shell of the foreign atom is actually ionized, and we have seen in § 3 that this is probably not the case. Thus, so far as the resistance due to impurities is concerned, nickel and palladium behave like copper or silver or gold.

On the other hand, this argument cannot apply to alloys containing comparable amounts of the two metals, because here ΔV must be taken to be the difference between the potential in the real lattice and the periodic potential which approaches most nearly to it(31), and so will be finite in the nickel or palladium atoms as well as those of the noble metal. Hence scattering of the electrons from *s* to *d* states will occur, so that the resistance rises more steeply than it otherwise would for increasing concentration of the noble metal.

This is the probable reason for the surprising sharpness of the maximum of the curve (III) of figure 5 as compared with that for silver-gold, palladium-platinum, etc.

A quantitative theory may be given as follows. If $V_A(r)$ is the potential energy of an electron in a palladium atom, and $V_B(r)$ in an atom of copper, silver or gold, and if x is the concentration of the noble metal, then, according to Nordheim, the periodic potential which approaches most nearly to the true potential is

$$V=(1-x)\,V_A+x\,V_B.$$

The difference between this and the potential in a palladium atom is

$$V-V_A=x\,\Delta V, \quad \Delta V=V_B-V_A,$$

and since there are $1-x$ of them, the probability of being scattered by a palladium atom is proportional to

$$(1-x)\,x^2\,|\int \psi_{\mathbf{k}'}\,\Delta V\,\psi_{\mathbf{k}}\,d\tau\,|^2.$$

Similarly, the probability of being scattered by a silver atom is proportional to

$$x\,(1-x)^2\,|\int \psi_{\mathbf{k}'}\,\Delta V\,\psi_{\mathbf{k}}\,d\tau\,|^2.$$

If now the *d* shells are full, we have simply to add these two terms, whence we see that the resistance is

$$\beta\,\{(1-x)\,x^2+x\,(1-x)^2\}=\beta x\,(1-x) \qquad \ldots\ldots(13),$$

β where β is a constant, which is Nordheim's result.

If, on the other hand, the palladium atoms have incomplete *d* shells, the probability of scattering by a palladium atom is greater than by a silver atom. Assuming in rough agreement with the experimental results shown in figures 4 and 5, that for the d band $N(E_{\text{max.}})$ is proportional to

$$(p-x)^2 \qquad x<p,$$

and zero otherwise, we see that the resistance due to transitions of the electrons into the d shells is

$$\alpha\,(p-x)^2\,x^2\,(1-x) \qquad \text{......(14),}$$

where α is another constant.

In figure 6 we show the two terms which contribute to the resistance; curve (II) is calculated from equation (14); curve (I) has the general form of equation (13), but we have shifted the maximum slightly to the right to take account of the fact that palladium has fewer electrons than silver. It is seen that with suitable choice of the ratio $\alpha:\beta$, the general form of the experimental curve of figure 5 can be reproduced, curve (III).

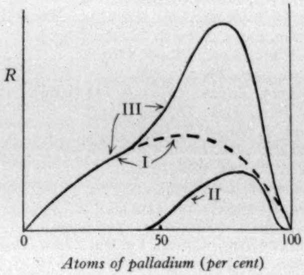


Figure 6. Theoretical curve for the resistance, at 0° K. of silver-palladium alloys: (I) due to s—s transitions; (II) due to s—d transitions; (III) total; (I) plus an arbitrary multiple of (II).

The constants α and β cannot be calculated without a detailed knowledge of the wave functions.

The experimental resistance curves show, however, that the effect of the d shells cannot increase the resistance by a factor of more than about two, as against a factor of five to ten for thermal agitation; this may be explained by the fact that ΔV is spherically symmetrical in the neighbourhood of any atom, and so formula (12) may be written in the following form, which should be compared with equation (10):

$$\left| C\int \psi_d^{n-1}\,\Delta V\,\psi_d^{n}\,d\tau \right|^2 \qquad \text{......(15),}$$

whereas the transition probability from one s state to another is

$$\left| A^2\int \psi_s^{n}\,\Delta V\,\psi_s^{n}\,d\tau \right|^2 \qquad \text{......(16).}$$

Since $C \ll A$ we may assume the quantity (15) to be small compared with (16).

§7. ACKNOWLEDGMENTS

In conclusion, I would like to express my thanks to Dr H. H. Potter, who has brought to my notice much of the experimental material discussed in this paper, and who has read it in manuscript and made several suggestions for its improvement.

REFERENCES

(1) Wigner and Seitz. *Phys. Rev.* **43**, 804 (1933); **46**, 509 (1934).
(2) Slater. *Phys. Rev.* **36**, 57 (1931).
(3) Jones, H. *Proc. roy. Soc.* **147**, 396 (1934).
(4) Sommerfeld and Bethe. *Handb. der Phys.* **24** (2), 428 (1933).
(5) Veldkamp. *Physica*, 's Grav., **2**, 15 (1935).
(6) Slater. *Phys. Rev.* **45**, 794 (1934).
(7) Grimm and Wolff. *Handb. der Phys.* **24** (2), 1073 (1934).
(8) Stoner. *Magnetism and Matter*, p. 381 (1934).
(9) Bloch. *Z. Phys.* **57**, 545 (1929).
(10) Wigner. *Phys. Rev.* **46**, 1002 (1934).
(11) Alder. *Promotionsarbeit*, Zürich (1916).
(12) Sadron. *Ann. Phys.*, Lpz., **17**, 371 (1932), and Thèse, Strasbourg (1932).
(13) Constant. *Phys. Rev.* **36**, 1659 (1930).
(14) Sucksmith. *Proc. roy. Soc.* A (in press).
(15) Pauli. *Z. Phys.* **41**, 81 (1926).
(16) Peierls, R. *Ergebn. exakt. Naturw.* **11**, 264 (1932).
(17) Svensson. *Ann. Phys.*, Lpz., **14**, 699 (1932).
(18) Vogt. *Ann. Phys.*, Lpz., **14**, 1 (1932).
(19) Aharoni and Simon. *Z. phys. Chem.* **4**, 175 (1929).
(20) Svensson. *Ann. Phys.*, Lpz., **18**, 302 (1933).
(21) Dorfmann. *Phys. Z. Sowjet.* **3**, 399 (1933).
(22) Linde and Borelius. *Ann. Phys.*, Paris, **84**, 747 (1927).
(23) Mott. *Proc. phys. Soc.* **46**, 680 (1934).
(24) Slater. *Phys. Rev.* **46**, 1002 (1934).
(25) Giebel. *Z. anorg. Chem.* **70**, 240 (1911).
(26) Borelius. *Handb. der Metallphysik*, p. 341 (1934).
(27) de Haas and Krupkowski. Comm. Leiden, **194** *a* (1930).
(28) Linde. *Ann. Phys.*, Lpz., **15**, 219 (1932).
(29) Norbury. *Trans. Faraday Soc.* **16**, 570 (1921) (review).
(30) Mott and Zener. *Proc. Camb. phil. Soc.* **30**, 249 (1934).
(31) Nordheim. *Ann. Phys.*, Lpz., **9**, 641 (1931).
(32) Keesom and Clark. *Physica*, 's Grav., **2**, 230 (1935).

55.037.2

EXPERIMENTS RELATING TO THE DISTRIBUTION OF ALTERNATING ELECTRIC CURRENTS IN THE EARTH AND THE MEASUREMENT OF THE RESISTIVITY OF THE EARTH*

By S. WHITEHEAD, M.A., Ph.D., F.Inst.P., A.M.I.E.E.

AND

W. G. RADLEY, Ph.D., M.I.E.E.

Received October 25, 1934. *Read in title May* 3, 1935.

ABSTRACT. The present memorandum discusses methods of determining the electrical resistivity of the earth, particularly from the point of view of the correlation of a.-c. and d.-c. methods. Various experimental methods are discussed together with the theory of current flow in homogeneous and stratified media. The topics considered are illustrated from experiments made to confirm the Carson-Pollaczek formulae for alternating currents in the earth and from surveys made from time to time in connexion with the problem of telephone interference. Various alternative methods are recommended for use in different circumstances.

§ 1. INTRODUCTION

A number of problems in electrical engineering, one of the most important being telephone interference(11), require a knowledge of the distribution of an electric current, generally alternating, in the earth. This depends among other things on the electrical resistivity of the earth. The authors have accordingly been led to experimental investigation of various methods of determining this resistivity. Within certain limits it has been found possible to correlate values obtained respectively by a.-c. and by d.-c. methods, which had previously proved a difficulty. Although the accuracy is adequate for the practical problems involved, the more fundamental aspects can be treated with only a relatively low degree of accuracy. The methods employed are described in this paper, together with field experiments which help to establish the Carson-Pollaczek theory for the distribution of an alternating current in a homogeneous earth.

§ 2. HOMOGENEOUS CONDITIONS

Alternating currents with parallel flow. When an alternating current flowing in a line returns through the earth, the distribution of the return current depends largely on its inductive effects, as does the skin effect in large metallic conductors.

---

* Report *M/T* 31 of the British Electrical and Allied Industries Research Association entitled "Experiments relating to the Distribution of Electric Currents in the Earth".

Maxwell's electromagnetic equations are applicable to the problem in the following steady-state form:

$$\left.\begin{aligned}\text{curl } H &= 4\pi\sigma\overline{E}\\ \text{curl } E &= -j\omega\overline{H}\end{aligned}\right\} \qquad \ldots\ldots(1),$$

where σ is the conductivity of the medium. From the solution of equations (1), a group of expressions is obtained which may be related to various eddy-current and skin-effect phenomena, as has previously been shown by one of the present authors(1). In practice, when interference between electric power and communication circuits is concerned, the problem usually resolves itself into a determination of the mean axial electric intensity along the communication line due to a given earth-return current flowing in the power line. Theoretical investigation of this problem is simplified by three assumptions, namely (i) that the earth is a semi-infinite homogeneous mass of uniform resistivity; (ii) that the current-flow in the earth is everywhere parallel to the disturbing line; and (iii) that the current in this line is constant throughout the section under consideration.

The second assumption neglects the crowding together of the lines of current-flow at the points where the earth connexions are made, and so limits the strict application of the resulting formulae to sections of a long straight line remote from its ends. The formulae are further restricted in their application by the fact that displacement currents in the earth have been neglected, but recent work indicates that the effect of the latter only begins to be apparent at frequencies above 60 kc./sec. In 1926, each starting with the above assumptions, F. Pollaczek(2) and J. R. Carson(3) independently derived formulae for the self and mutual impedance of parallel earthed wires. The two sets of formulae may be regarded as limiting cases of the solution of equations (1), in terms of cylindrical and rectangular coordinates respectively. Their equivalence corresponds to the expression of Bessel functions as Fourier integrals. For a given set of conditions, Pollaczek's and Carson's formulae give numerical results which are identical.

The coupling between two parallel lines, being partly resistive and partly reactive, is best expressed in terms of a mutual impedance. This is the practice in America, but in Europe the coupling has always been given as a generalized coefficient of mutual induction, thus

$$M = -\frac{4j}{\gamma^2} + 4\,\frac{\text{kei}'\,\gamma - j\,\text{ker}'\,\gamma}{\gamma} \qquad \ldots\ldots(2),^*$$

M where M is the generalized coefficient of mutual induction per cm. between the inducing and the induced line in c.g.s. units;

γ $\gamma = x\sqrt{(4\pi\sigma\omega)}$;

ω, f ω is 2π times the frequency f of the inducing current;

σ σ is the conductivity of the earth in c.g.s. units; and

x x is the separation of the two circuits in centimetres.

---

* Tables of ker′ and kei′ functions are given with the *British Association Report* of 1915, pp. 36–38; also as table 1 of "Bessel Functions for A.-c. Problems", by H. B. Dwight, *Trans. Amer. Inst. elect. Engrs*, pp. 812–20 (1929).

Although expressed as an inductance, M is a complex quantity, and is in reality the total mutual impedance per unit length divided by $j\omega$.

Experiments were carried out on Shap Fells, Westmorland, with specially constructed lines and measuring apparatus contained in testing vans which had been loaned by the German Reichspost. The site was selected as being free from stray currents owing to its remoteness from other power systems, and also because there is a close and very long parallel in that neighbourhood between a Central Electricity Board 132-kV. transmission line and a main Post Office trunk route. The lay-out of the lines used for the tests is shown schematically in figure 1. A light V.I.R. insulated and braided cable (3/0·029) was used for the primary line, which was 5 km. long. At first laid on the ground, this was later, owing to variations in its measured impedance, raised on 10-ft. poles, when consistent results were obtained. A cable containing one twisted pair of conductors within an earthed lead sheath was used to connect the line with the generators. Table 1 shows the results of impedance-measurements on the primary line after poling. The d.-c. resistance of this line with its earth plates was about 105 Ω. Approximately 70 Ω. of this was in the line itself and the connexions to the testing point.

The secondary lines were 3 km. long and consisted of light army field cable laid on the ground, at separations of 10, 100, 300, 1000 and 2000 metres from the inducing line. The secondary lines were terminated by lead earth plates, but a long length of buried copper strip was used at each end of the primary line. Owing to the rocky nature of the ground special precautions were necessary in order to reduce the resistance of all the earth connexions, and considerable quantities of salt were used for this purpose.

In all the experiments the total mutual impedance Z between the lines was derived from a measurement of the induced voltage e on a secondary line for a given current I_0 in the primary line. The majority of the measurements were made with the Franke machine for the higher frequencies (2000 to 300 c./sec.) and a Larsen potentiometer for the lower frequencies (300 to $16\frac{2}{3}$ c./sec.). A complete series was also made with the Campbell potentiometer for all frequencies. The results obtained in this way agreed very well with those obtained by the other methods.

Z
e
I_0

The Franke machine is essentially a generator having two armatures, the e.m.fs. generated in which can be varied in magnitude and phase with respect to one another. The circuit is shown in figure 2. A current of about 1 A. was supplied to the primary line from the lower (phase) armature through an amplifier when necessary. For tests at the highest frequencies a series condenser, set to yield resonance with the inductance of the line at the test frequency, was used to increase the current in the primary line. The upper (amplitude) armature of the Franke machine was connected to a potentiometer, and the measurements were then made as follows. The resistance R, in series with the ammeter in the primary line being set at some suitable value, of the same order as Z, for each series of tests, the phase angle ϕ_1 between the two armatures and the setting of the amplitude armature were varied until the potential-difference across some convenient fixed value r_1

R
ϕ_1
r_1

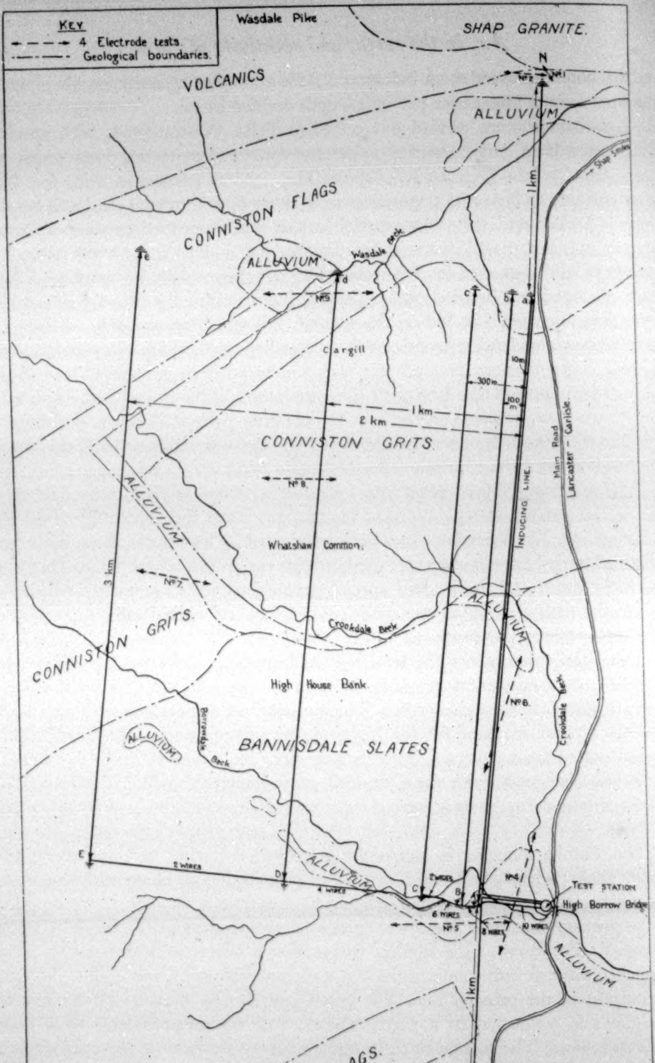

KEY
4 Electrode tests.
Geological boundaries.
Wasdale Pike
SHAP GRANITE.
N
Nº2
Nº1
VOLCANICS
ALLUVIUM
Shap 5 miles
1 km
CONNISTON FLAGS
e
ALLUVIUM
Wasdale Beck
d
Nº3
c
b
a
Cargill
10m
300m
100 m
1 km
2 km
CONNISTON GRITS.
Nº 8.
ALLUVIUM
INDUCING LINE.
Main Road
Lancaster - Carlisle.
Whatshaw Common.
3 km
Nº7.
Crookdale Beck
ALLUVIUM
CONNISTON GRITS.
High House Bank.
Nº6.
Crookdale Beck
Borrowdale Beck
ALLUVIUM
BANNISDALE SLATES
E
2 WIRES
D
4 WIRES
ALLUVIUM
C
2 WIRES
B
A
6 WIRES
Nº4
TEST STATION
High Borrow Bridge
Nº 5
8 WIRES
10 WIRES
ALLUVIUM
1 km
Nº3
S
KIRBY MOOR FLAGS.
Kendal 8 miles

Figure 1. Map of test site and lines at Shap.

on the potentiometer was equal and opposite to that across R. The potential-difference across the resistance R was then replaced by the induced voltage in the secondary line, and the fraction tapped off the potentiometer resistance r_2 and the phase angle ϕ_2 were varied until balance was obtained. Then r_2 ϕ_2

$$Z=(r_2/r_1)\,R \qquad \ldots\ldots(3),$$

since the current in the primary line and that in the potentiometer were the same for the two measurements. The phase angle between the induced voltage and the inducing current is given directly by $(\phi_2-\phi_1)$. A telephone was used to determine the conditions of balance, the filter shown in figure 2 being adjusted to cut off all harmonics of the test frequency.

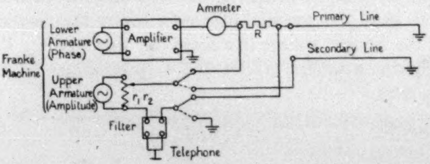

Ammeter
Primary Line
Lower Armature (Phase)
Amplifier
R
Secondary Line
Franke Machine
Upper Armature (Amplitude)
r_1 r_2
Filter
Telephone

Figure 2. Franke machine method used for measurements at high frequencies.

The Larsen potentiometer circuit is shown in figure 3. A measured current I_0 of about 1 A. was fed into the primary line from one of two small motor alternators having complementary frequency-ranges. As the secondary of the variometer was virtually on open circuit when balance was obtained, the voltage induced in it was 90° out of phase with respect to the current in the low-resistance slide wire. The induced voltage in the secondary line was balanced by varying the position of the slide-wire contact for the in-phase component, and by adjusting the variometer for the out-of-phase component. If r is then the slide-wire reading and m the mutual inductance between the two coils of the variometer I_0 r, m

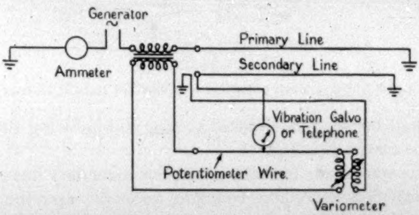

Generator
~
Ammeter
Primary Line
Secondary Line
Vibration Galvo or Telephone
Potentiometer Wire
Variometer

Figure 3. Larsen potentiometer used for measurements at low frequencies.

$$Z=n\sqrt{(r^2+\omega^2 m^2)} \qquad \ldots\ldots(4),$$

n being the known transformation ratio of the transformer in the primary line. ϕ is given by arc tan $\omega m/r$. A vibration galvanometer, tuned to the frequency of the inducing current, was used as a detector. n

The Campbell-Larsen potentiometer shown in figure 4 is a modification of the simple Larsen potentiometer just described. By incorporating the loop shunt shown diagrammatically, the current in the primary of the mutual inductometer can be varied with respect to the reference current in the resistance portion in such a way as to enable the out-of-phase component of the measured potential-difference to be read off directly in volts. This potentiometer has been found suitable for tests at all frequencies. Current at the lower frequencies was supplied to the primary line from a small alternator through the primary of a transformer and appropriately fixed resistance R. A tuned galvanometer was used as detector. For the higher frequencies the Franke machine was used as a generator. One armature (with amplifier if necessary) supplied current to the primary line, the other providing the reference current for the potentiometer. A telephone was then used to indicate balance. In both cases the potential-difference across the resistance R was first measured, then the induced voltage in a secondary line. If V_{r_1}, V_{m_1} and V_{r_2}, V_{m_2} are the potentiometer readings (1) for the voltage across R, and (2) for the induced voltage, then

$$Z = \frac{\sqrt{(V_{r_2}{}^2 + V_{m_2}{}^2)}}{\sqrt{(V_{r_1}{}^2 + V_{m_1}{}^2)}}\, R \qquad \text{......(5)},$$

R

V_{r_1}, V_{m_1}, V_{r_2}, V_{m_2}

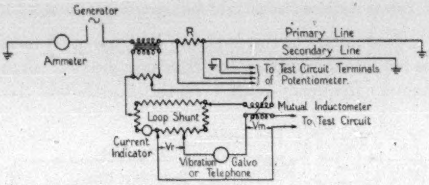


Figure 4. Campbell-Larsen potentiometer used for tests at all frequencies.

and the phase angle between the induced voltage and inducing current is equal to (arc tan V_{m_2}/V_{r_2} − arc tan V_{m_1}/V_{r_1}).

Measurements were made for each of the five secondary lines at the following frequencies: 2000, 1600, 1200, 800, 600, 400, 300, 200, 140, 100, 65, 45, 30 and 16⅔ c./sec. Repetition of the measurements at the lower frequencies gave results which agreed, but some difficulty was experienced in obtaining such agreement at frequencies between 600 and 2000 c./sec. It was shown later that the insulated cable used for the secondary lines had a large variable earth-capacity under changing weather conditions. This had a complex effect on the measurement owing to the joint effects of capacity leakage and capacity coupling. The mean values for a given frequency and separation were consistent, and are shown on figure 5.

By virtue of equation (2) the conductivity of the earth, assumed to be uniform, may be deduced from the coefficient of mutual induction. Measurements made at frequencies below 100 c./sec. being neglected, the results obtained from the secondary lines at 300, 1000 and 2000 metres separation (*Cc*, *Dd* and *Ee* on figure 1) are consistent with an earth conductivity of $4{\cdot}5 \times 10^{-15}$ c.g.s. u., i.e. a resistivity of 222,000 Ω.-cm. For the lines at 10 metres and 100 metres separation (*Aa* and

Bb) slightly higher values of the conductivity, equal respectively to $6{\cdot}5 \times 10^{-15}$ and $5{\cdot}5 \times 10^{-15}$ c.g.s. u. (corresponding to resistivities of 150,000–180,000 Ω.-cm.), would have given better agreement with the observations. At close separations, however, variation in the earth-conductivity does not so greatly affect the coupling between two lines, and the value of the conductivity may be altered without causing much change in the theoretical value of the coefficient of mutual induction.

Assuming a uniform conductivity equal to $4{\cdot}5 \times 10^{-15}$ c.g.s. u. for all the lines, the theoretical variation of the mutual inductance with frequency is shown on figure 5. For all the lines there is fairly close agreement between the measured values and the calculated curve from 2000 to 100 c./sec. At lower frequencies the observed values were greater than those given by the theory. This was due to the primary line being of finite length, and is discussed later.

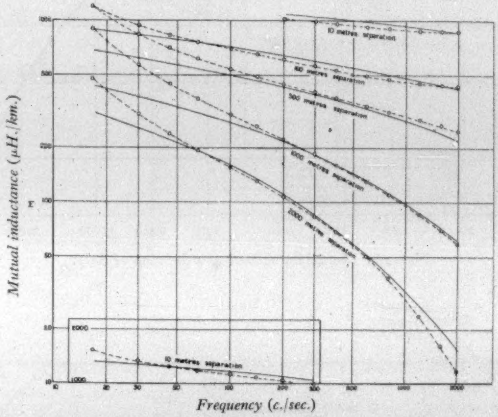


Figure 5. Variation of *M* (mutual inductance) with frequency. ○ Measured value; --○-- experimental curve; ——— calculated curve for $\sigma = 4{\cdot}5 \times 10^{-15}$ c.g.s.

Referring to equation (2), it is observed that the mutual inductance is a function of the separation multiplied by the square root of the frequency for a given uniform earth-resistivity; so that, if plotted against this parameter, all experimental results should fall on one curve. That this was largely the case at Shap (apart from the correction due to the finite length of the inducing line) is shown by figure 6. It may be taken as a criterion of current-flow according to the Carson-Pollaczek theory.

A further verification of the theory was obtained at Eltham where the earth is of low resistivity. Measurements were made with the Campbell-Larsen potentiometer circuit, already described, and with the lay-out shown on figure 7. In this case, however, the current was led into the earth from an underground cable and it was necessary to correct for the part of the current that returned in the sheath of the cable and by adjoining buried conductors. The fraction varied with frequency and was determined experimentally by supplying current to the outer secondary line and comparing the induction from this with that from the cable. It was found that when the screening effect of the cable sheath was allowed for, the results from both sets of experiments were consistent with one another and with theory, and with a conductivity of the earth equal to $1{\cdot}5 \times 10^{-13}$ c.g.s. u. or a resistivity of 667 Ω.-cm. The induction from the cable corrected for the effect of the sheath is shown by figure 8.

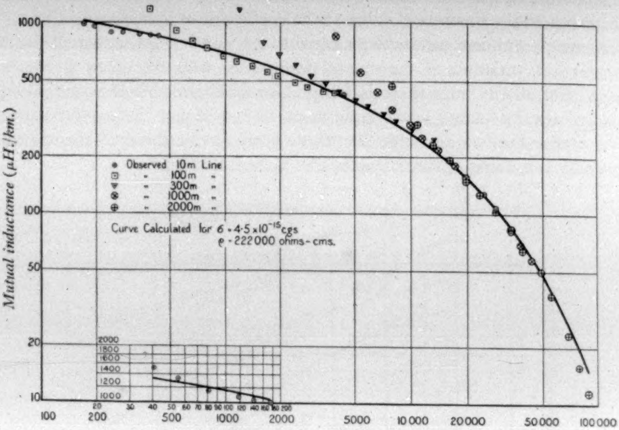


Figure 6. Mutual inductance as a function of $S\sqrt{f}$.

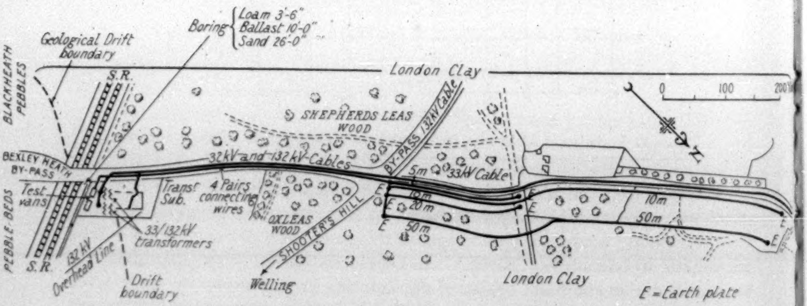

Boring { Loam 3'-6" Ballast 10'-0" Sand 26'-0"
Geological Drift boundary
BLACKHEATH PEBBLES
S.R.
London Clay
0
100
200
SHEPHERDS LEAS WOOD
BY-PASS 132kV Cable
BEXLEY HEATH BY-PASS
32kV and 132kV Cables
5m
33kV Cable
Test vans
Transf Sub.
4 Pairs connecting wires
E
10m
20m
50m
PEBBLE-BEDS
OXLEAS WOOD
SHOOTER'S HILL
33/132kV transformers
S.R.
132 kV Overhead Line
Drift boundary
Welling
London Clay
E = Earth plate

Figure 7. Map of test site and lines at Eltham.

So far, attention has been confined to the longitudinal e.m.f., that is, the e.m.f.

induced in a long straight conductor parallel to the inducing line. In addition to this, a search coil placed on the ground will have a transverse e.m.f. induced in it owing to the vertical component of the magnetic field. The transverse e.m.f. is proportional to dM/dx and if measured at different frequencies and separations will enable the resistivity to be deduced. Collard has developed this method and a publication describing it is anticipated. Collard obtained by this method a value of 1500 Ω.-cm. by displacements along the Shooter's Hill By-Pass Road, figure 7, and a value of 1000 Ω.-cm. at the north-west end of the test section by displacements normal to the cable.

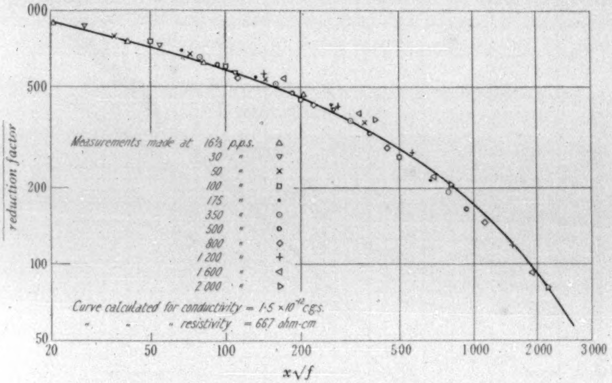


Figure 8. Mutual induction as a function of $x\sqrt{f}$.

Both the longitudinal and transverse methods can be and have been used to measure an unknown earth current of any suitable frequency by the variation of e.m.f. with separation from the source.

Alternating currents with divergent flow. The lines of flow of an alternating current between two point electrodes is roughly indicated in figure 9. Between the broken lines is the region of parallel flow already considered. Outside the broken lines the flow is divergent and usually complicated. The problem has been treated by a number of authors and a solution by Foster⁽⁴⁾ in terms of Fourier integrations may be consulted. In practice, however, lack of uniformity of the

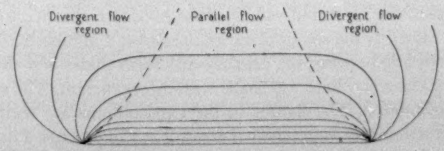


Figure 9. End effects with alternating current.

conditions has usually rendered the application of general theory uncertain. Certain approximations may sometimes be made.

Referring to the tests carried out at Shap with the frequency of the inducing current decreased below 100 c./sec., the experimentally-measured total mutual impedance was divided into its resistive and reactive components, as indicated by the measured value of its angle. The resistance components have been plotted on figure 10. In the same figure the theoretical values of this component have been plotted for each of the secondary lines, an infinite primary line and a constant value of the earth conductivity equal to $4{\cdot}5 \times 10^{-15}$ c.g.s.u. being assumed.

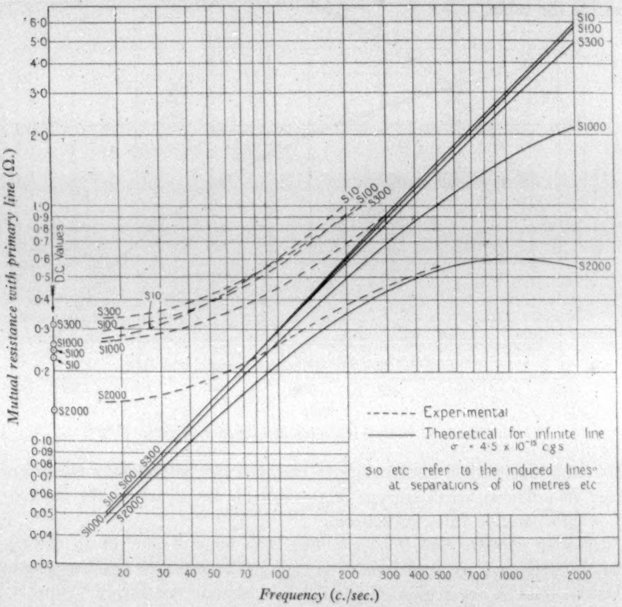


Figure 10. Mutual resistance of lines at Shap—variation with frequency.

At $16\frac{2}{3}$ c./sec., the measured values of the mutual resistance were in excess of the calculated values by the amounts shown in column 2 of table 2. Now a measurement of the coupling between the primary-line earth plates and those terminating each of the five secondary lines was made with direct current. The results are given in column 3 of table 2. The figures in the two columns are seen to correspond almost within the limits of experimental error. Thus in practice a limit to the error, introduced by divergent flow at the ends of the line, can be fixed experimentally.

Alternatively the mutual resistance R between the two sets of earth plates could have been calculated from the Wenner formula[(5)], thus

R

$$R = \frac{\rho}{2\pi}\left[\frac{1}{a_{13}} + \frac{1}{a_{24}} - \frac{1}{a_{14}} - \frac{1}{a_{23}}\right] \qquad \text{......(6)},$$

where current flows in the earth between points 1 and 2 and the potential difference is measured between 3 and 4. a_{mn} is the distance between two points m and n, and ρ the resistivity of the earth assumed to be isotropic and homogeneous. Actually it will be noted from table 2 that this last condition did not hold exactly for the test lines at Shap, where owing to some geological irregularity the earth-plate coupling between the primary line and s_{300} was greater than that for either of the two nearer lines. Equation (6) applies strictly to direct current only, but can be used for alternating current in the absence of skin effect. Where skin effect can be avoided, the freedom of alternating current from troubles due to polarization and spurious ground potentials has led to its frequent use. Generally the problem is to determine ρ from a measurement of the mutual resistance between two pairs of electrodes. For convenience the four electrodes are uniformly spaced in a straight line, 1 and 2 being at the ends of the line. If a is the spacing between adjacent electrodes, equation (6) reduces to

a_{mn}

ρ

a

$$\rho = 2\pi aR \qquad \text{......(7)}.$$

Table 1. Impedance-measurements made on the inducing line.
Length of line, 5 km.

Frequency	Impedance		L (mH./km.)	R total (Ω.)
	Ω.	ϕ		
2000	239	54° 35′	3·11	138·6
1600	204	50° 25′	3·12	129·6
1200	165	43° 50′	3·03	119
800	134	33° 35′	2·94	111·4
600	122·5	26° 40′	2·91	109·5
400	113·2	19° 00′	2·93	107
300	111·4	14° 15′	2·91	108
200	109	9° 45′	2·94	107·4

Table 2. Mutual resistance of earth-plate system.

Line	Mutual resistance (Ω.)	
	Excess measured over calculated value at 16⅔ c./sec.	Measured d.-c. coupling
s_{10}	0·22	0·235
s_{100}	0·24	0·250
s_{300}	0·28	0·310
s_{2000}	0·215	0·255
s_{3000}	0·10	0·135

The general features of the method based on equation (7) will be considered later in connexion with d.-c. measurements, but the conditions for avoiding a.-c. difficulties are important. Skin effect is reduced by decrease of electrode spacing, decrease of frequency, and increase of resistivity, but it will introduce a large error at most practicable frequencies if R is less than $0{\cdot}1\ \Omega$. Owing to high electrode-resistance it is necessary to use a null method or a high-impedance circuit to determine the potential between the points 3 and 4. A.-c. potentiometers are usually expensive and rarely suitable for the type of field work required. A simplified potentiometer circuit which has been employed is shown in figure 11. At low frequencies the detector used in the potentiometer circuit offers difficulties. If it is non-synchronous and non-linear (for instance of the metal rectifier type) it will be insensitive and subject to stray alternating currents. If the detector is tuned, as in the case of a vibration galvanometer, portability is difficult. Also the balance will be imperfect owing to phase-displacements. If linearity is obtained by the addition of an alternating e.m.f., as in bridge and valve methods, a phase-changing device is usually necessary. In various early investigations, made by the Electrical Research Association, in which the stray a.-c. was serious, it was necessary to use a heavy alternating current and to measure the potentials with a valve voltmeter having an input impedance of 0·1 megohm.

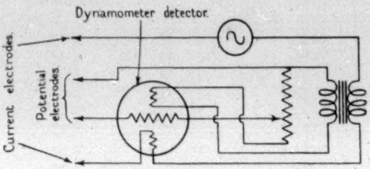


Figure 11. A.-c. earth tester—null method.

An apparatus of greater portability in which an amplifier valve voltmeter is used as a transfer standard has since been found satisfactory in eliminating the effect of stray alternating and continuous e.m.fs. The current electrodes are supplied with current at 30 c./sec., through a fixed series resistance having a variable potentiometer across it. The amplifier valve voltmeter is switched from the potential electrodes to this potentiometer, which is adjusted to obtain equality of reading. The mutual resistance in ohms may then be read from the potentiometer. The amplifier stages are coupled by low-pass filters so that e.m.fs. of frequencies of 50 c./sec. and upwards are eliminated. They can be several times as great as the test e.m.f. of 30 c./sec. without error. The instrument can be used for e.m.fs. greater than 1 mV. and is designed for current-ranges of 1 and 10 A., giving ample sensitivity. It is shown schematically on figure 12, but a fuller description will be published later. It is possible to rectify part of the main current and use the equipment as a null instrument, but certain difficulties as regards permanence and accuracy, which have not yet been finally investigated, arise under these conditions.

Stray d.-c. and polarization e.m.fs. are avoided by means of a condenser input to the valve voltmeter.

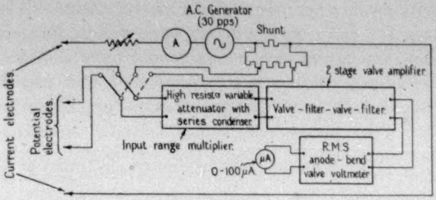

A.C. Generator
(30 pps)
Shunt
A
2 stage valve amplifier
High resistor variable attenuator with series condenser
Valve - filter - valve - filter
Input range multiplier
R.M.S anode - bend valve voltmeter
μA
0-100μA
Current electrodes
Potential electrodes

Figure 12. Electrical Research Association earth tester (a.-c.) for use in tuned transfer method.

The handy and portable megger earth tester is well known. This utilizes the ohm-meter principle and measures the mutual resistance directly. A more sensitive instrument, also employing commutation and rectification but utilizing a potentiometer and galvanometer in the measuring circuit, is now on the market.

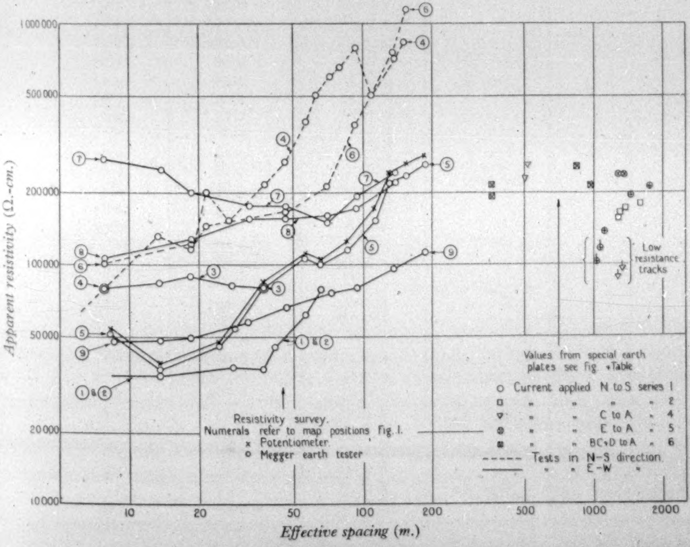


Figure 13. Four-electrode tests at Shap.

Direct currents. Skin effect is absent with direct current so that attention is confined to end effects. The same principles apply as with alternating current under conditions such that the skin-effect region is negligible. It is again convenient to consider the earth as a four-terminal resistance with two current electrodes 1 and 2 and two potential electrodes 3 and 4, and equation (6) may be used to determine the resistivity ρ from a mutual-resistance measurement.

If the points 1, 2, 3 and 4 are not on the same straight line, errors are greater and topographical peculiarities have a greater and more complex effect. In the Shap experiments already referred to, the special lines were used for determining the mutual resistances of various groups of earthing-points and the resistivity deduced therefrom. Table 3 shows some of the results, the designation of the current supply and potential points being shown in figure 1. On the same site a resistivity-survey was made with electrodes equally spaced in a straight line and with either a megger earth tester or d.-c. potentiometer. The results obtained have been plotted against electrode-separation on figure 13. Some of the values given in table 3 have also been inserted on figure 13, the effective spacing in this case being

Table 3. D.-c. measurements on special electrodes at Shap

Series	Current supply	Potential-measurement	Mutual resistance (Ω.)	Resistivity (Ω.-cm.) × 10⁵
1	*S* to *N*	*A* to *a*	0·235	1·0
		B to *b*	0·250	1·15
		C to *c*	0·310	1·35
		D to *d*	0·255	1·85
2	*S* to *N*	*E* to *e*	0·135	2·05
		e to *A*	−0·255	1·6
		e to *B*	−0·255	1·65
		e to *C*	−0·245	1·7
		e to *D*	−0·19	1·75
3	*S* to *A*	*e* to *B*	3·46	2·15
		e to *C*	1·03	2·65
		e to *D*	0·15	3·7
		e to *E*	0·06	—
4	*e* to *A*	*E* to *D*	0·14	1·65
		E to *C*	1·20	2·5
		E to *B*	3·61	2·1
		d to *c*	0·39	0·85
		d to *b*	1·57	0·95
		d to *a*	−3·10	32
		d to *B*	3·78	2·05
		c to *C*	1·205	2·2
5	*E* to *A*	*D* to *C*	1·035	2·25
		D to *B*	3·87	2·25
		e to *d*	0·09	—
		e to *c*	0·94	—
		e to *b*	0·105	—
		e to *a*	0·165	—
6	*D* to *A*	*E* to *C*	1·00	2·5
		E to *B*	3·43	2·05
	C to *A*	*E* to *D*	−0·09	1·6
		E to *B*	1·90	1·9
	B to *A*	*E* to *D*	−0·03	—
		E to *C*	−0·52	2·1

taken as the geometric mean of the electrode-intervals. In view of the type of geological structure, there is a reasonable measure of correlation. The location of the straight-line survey measurements can be seen from the map of the site, figure 1.

The more important difficulties inherent in the measurements are as follows. (i) All four contacts of the four-electrode resistance formed by the earth are of high and sometimes variable resistance; (ii) a potential-distribution in the earth may arise from natural causes; (iii) stray and sometimes rapidly varying e.m.fs. may arise from neighbouring d.-c. supply or electric traction systems; (iv) polarization of the potential electrodes may cause spurious potential-readings. It is not easy in one method to overcome all these difficulties. The Kelvin double bridge overcomes (i), but must be of excessive resistance in the secondary circuits, and is precluded by the remaining difficulties and by the magnitude of supply voltages used, so that a potentiometer or high-resistance galvanometer has generally to be used as a voltmeter. The e.m.fs. under (ii) are usually steady and may be compensated with an added e.m.f., or an initial zero reading may be taken. Unless the potential-difference between the potential electrodes is made fairly large, non-polarizable electrodes must be used to overcome the difficulties under (iv). A form used by the authors, similar to that designed by Broughton-Edge, may be quickly constructed from easily obtainable components; it is shown in figure 14. The porous pot is of the type used in primary cells. The protruding stem is made long enough for the connexion, which may be a clip, socket or terminal, to be kept dry and clean as far as possible. Stray leakage e.m.fs. under (iii) can only be overcome by the use of sufficiently heavy main currents or by waiting for the occurrence of steady periods. In the latter connexion the damping of the galvanometer employed should be as nearly critical as possible subject to the period not exceeding the order of 1 or 2 sec. A double suspension or semi-suspended type

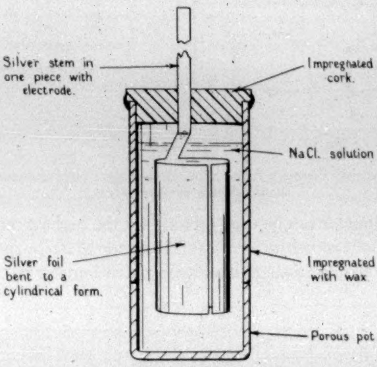

Silver stem in one piece with electrode.
Impregnated cork.
NaCl. solution
Silver foil bent to a cylindrical form.
Impregnated with wax
Porous pot

Figure 14. Simple form of non-polarizable electrode.

is suitable. There are several types of potentiometer available, for both special and general purposes. A type used by the Electrical Research Association is shown in figure 15. It has the advantage of a wide range, the lowest range being 0–300 μV., while the 3-volt voltmeter can be used directly for the highest range. The accuracy, although only that of the voltmeter, is sufficient for these purposes. The contacts are fixed, moving contacts causing difficulty in potentiometer circuits for field work.

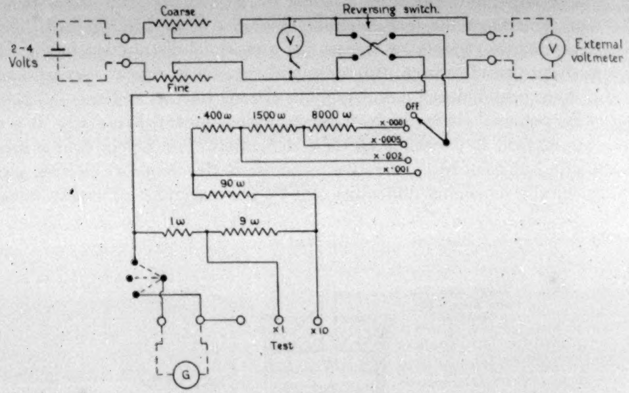
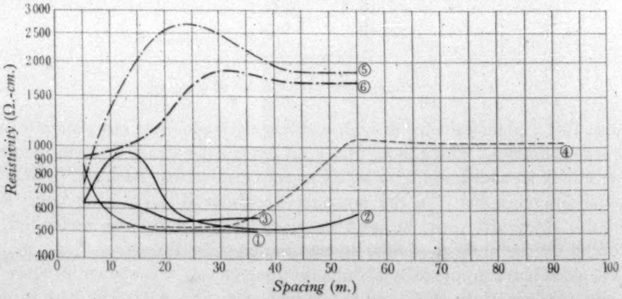


Figure 15. Electrical Research Association fixed-contact potentiometer for field and thermoelectric measurements.



—— (1) Test site parallel to cable. —— (2) Test site 45° to cable. —— (3) Test site perpendicular to cable. ---- (4) Along Shooter's Hill By-Pass Road. –·–· (5) Adjacent to substation parallel to cable. –·–· (6) Adjacent to substation perpendicular to cable.

Figure 16. Four-electrode tests at Eltham.

Figure 16 shows the result of a resistivity-survey at Eltham made with a d.-c. potentiometer. The site, shown on figure 7, was of unusual difficulty owing to the very low resistivity and the presence of an electric railway. Resistivity-values deduced from a.-c. mutual-inductance and search-coil measurements have already been given and on the whole the agreement is satisfactory.

Figure 17, prepared from the geological survey map, shows in vertical section the approximate geological structure along a route between Carlisle and Lancaster. Measurements of the resistivity were made at various points along this route by the megger earth tester and a.-c. search-coil methods. Values obtained from the two methods agreed with one another and with the geological formation. They also agreed fairly well with measurements of the induction from a power transmission line at various points along the route.

The examples which have been given in this section of the paper illustrate (i) the order of agreement between different methods which may be obtained even under unfavourable conditions; (ii) the difficult condition produced by a structure which, though of high resistivity, is folded and broken with a very uneven contour. We observe that some d.-c. results in the north-to-south direction at Shap gave values suggesting a lower insulating layer which is not observed with a.-c., whereas the d.-c. results in the east-to-west direction are more consistent. This may be due to local faults and cleavage planes which have a low a.-c. impedance but high d.-c. resistance; (iii) the difficulties associated with a low resistivity and extraneous disturbing currents.

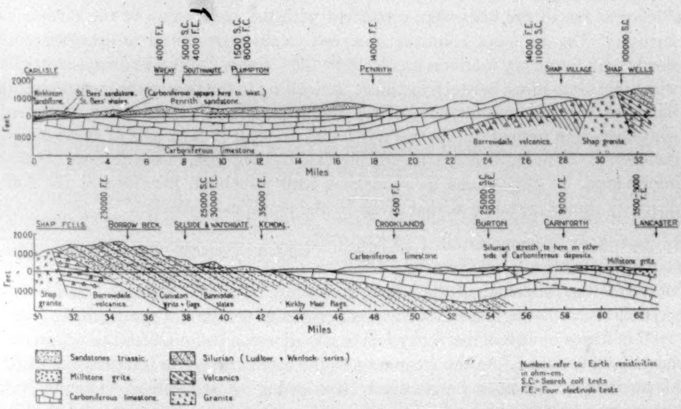

4000 F.E.
5000 S.C.
4000 F.E.
1500 S.C.
8000 F.C.
14000 F.E.
14000 F.E.
11000 S.C.
100000 S.C.
CARLISLE
WREAY
SOUTHWAITE.
PLUMPTON
PENRITH
SHAP VILLAGE.
SHAP WELLS
2000
1000
0
1000
Feet
Kirklinton sandstone.
St. Bees' sandstone.
St. Bees' shales.
(Carboniferous appears here to west.)
Penrith sandstone.
Carboniferous limestone
Borrowdale volcanics.
Shap granite.
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
Miles
230000 F.E.
25000 S.C.
30000 F.E.
35000 F.E.
4500 F.E.
25000 S.C.
30000 F.E.
9000 F.E.
3500-9000 F.E.
SHAP FELLS.
BORROW BECK.
SELSIDE & WATCHGATE.
KENDAL.
CROOKLANDS
BURTON
CARNFORTH
LANCASTER
2000
1000
0
1000
Feet
Shap granite.
Borrowdale volcanics.
Coniston grits + flags.
Bannisdale slates.
Kirkby Moor flags.
Carboniferous limestone.
Silurian stretch to here on either side of Carboniferous deposits.
Millstone grits.
30
32
34
36
38
40
42
44
46
48
50
52
54
56
58
60
62
Miles
Sandstones triassic.
Millstone grits.
Carboniferous limestone.
Silurian (Ludlow + Wenlock series)
Volcanics
Granite
Numbers refer to Earth resistivities in ohm-cm.
S.C.= Search coil tests
F.E.= Four electrode tests

Figure 17. Estimated rock section: Lancaster-Carlisle test route.

It has been further confirmed by the authors that measurements on outcrops offer no clue to the resistivity of strata. Electrodes were inserted in pneumatic drill holes in granite and other stony outcrops, but the results could not be correlated with the main observations. Measurements of the electrode-resistance of cylindrical electrodes by the three-electrode method agreed with four-electrode tests, but the accuracy with which the resistivity can be deduced therefrom is very much less.

§ 3. NON-HOMOGENEOUS CONDITIONS

Alternating currents with parallel flow. The German Reichspost co-operated with the Commission Mixte Internationale* to carry out two extensive series of field experiments in 1928[(6)]. One of these was in the Oldenburg fen district, the other in a limestone district at Munsingen in Würtemberg. In both places test lines, 5 km. long, were specially constructed at different distances apart up to 3 km., and use was made of the specially equipped testing-vans which were later loaned for the Shap experiments.

At any one given frequency of the testing-current the observed mutual inductances for all the lines were consistent with the same value of the earth-conductivity. On the other hand the observed values were found to decrease more slowly with frequency than was predicted by the Carson-Pollaczek theory, especially with large separations between the lines. According to the theory the measurements indicated that the conductivity of the earth, assumed homogeneous, decreased with increasing frequency. As this was found to occur in both sets of experiments, the Comité Consultatif International† directed in 1929 that, in the absence of other information, in calculations in connexion with telephone interference the conductivity should be taken as that given by the empirical formula

$$\sigma = \frac{1 \cdot 5 \times 10^{-12}}{\sqrt{f}} \text{ c.g.s.u.} \qquad \ldots\ldots(8),$$

f where f is the frequency of the inducing current. The coefficients in this formula were chosen to give agreement with the experimental results at Munsingen, figure 18.

It will now be shown that a physical explanation can be found for this apparently anomalous variation. As the frequency of the earth current is increased its depth of penetration becomes progressively less owing to skin effect. Therefore if, although regarded as homogeneous, the conductivity of the earth in reality changes with depth below the surface, its mean conductivity will appear to vary with frequency.

A recently published note from H. J. Josephs[(7)], a colleague of one of the authors, extends the Carson-Pollaczek formula, equation (4), to the more general case of the mutual impedance between two wires on the surface of a stratified earth.
b It is assumed that the earth consists of an upper layer of depth b and uniform
σ_1 conductivity σ_1. Below this layer the earth is supposed to have a uniform con-
σ_2 ductivity σ_2. Neglecting end effects the mutual impedance per unit length between the two wires is given to a close approximation by

$$Z = 4\omega \{(P + P_s) + j(Q + Q_s)\} \qquad \ldots\ldots(9).$$

P, Q P and Q are one quarter the *imaginary* and *real* parts respectively of the right-hand side of equation (2) and are obtained on the assumption that the conductivity σ_1

---

* Commission Mixte Internationale pour les Expériences rélatives à la Protection des Lignes de Télécommunication et des Canalisations Souterraines.

† Comité Consultatif International des Communications Téléphoniques à Grande Distance.

extends uniformly downwards without interruption. The terms P_s and Q_s correct for the change in conductivity from σ_1 to σ_2 at the depth b,* and are given by

P_s, Q_s
b

$$\left.\begin{aligned} P_s &= \int_0^{2\pi} N_s\, e^{-g\sqrt{\beta}} \cos g\gamma\, dg \\ Q_s &= -\int_0^{2\pi} N'_s\, e^{-g\sqrt{\beta}} \cos g\gamma\, dg \end{aligned}\right\} \quad \text{......(10),}$$

where N'_s and N_s are the real and imaginary parts respectively of

N'_s, N_s

$$\left[\frac{\alpha}{(g+\alpha)^2}\,\frac{2U}{U-1}\right],$$

and

$$\sqrt{\beta} = \sqrt{(4\pi\sigma_1\omega)}$$

β

$$\alpha = \sqrt{(g^2+j)}, \quad \phi = \sqrt{(g^2+jK)}$$

α

$$K = \sigma_2/\sigma_1$$

$$U = \frac{(\phi-\alpha)^2}{j(K-1)}\, e^{-2b\alpha\sqrt{\beta}},$$

U

and the other symbols have the meaning assigned to them in equation (4) with the conductivity taken as that of the uppermost stratum.

Similar expressions for the mutual impedance of earthed wires in the case of a horizontally stratified two-layer earth, which include the effects due to divergent flow at the terminations, have been published by J. Riordan and E. R. Sunde$^{(8)}$.† These cannot readily be evaluated in a closed form and, for numerical calculation of the mutual impedance with a stratified earth, methods based upon graphical integration become necessary, whether these expressions or those due to Josephs are used. In this way the curves shown in figures 18 and 19 have been calculated. The values of the three variables σ_1, σ_2 and b were chosen by trial to give agreement between the calculated curves and the values of the mutual inductance observed at Munsingen and Oldenburg respectively. Considerable assistance, however, was obtained from a scrutiny of the manner in which the conductivity, calculated on the assumption of a homogeneous earth, varied with frequency. The values chosen were as follows:

For Munsingen, $\sigma_1 = 40 \times 10^{-15}$, $\sigma_2 = 400 \times 10^{-15}$ c.g.s.u.
$b = 250$ metres.

For Oldenburg, $\sigma_1 = 80 \times 10^{-15}$, $\sigma_2 = 800 \times 10^{-15}$ c.g.s.u.
$b = 200$ metres.

In the case of Oldenburg, the agreement could be improved by assuming a rather greater depth b for the plane of stratification, with a corresponding increase in the conductivity of the lower layer, in order to annul the effect of this change on mutual inductances calculated for the lower frequencies. At the higher frequencies

---

* When $\sigma_1 = \sigma_2$, $K = 1$, and $P_s = Q_s = 0$.

† M. Gray$^{(8)}$ has also given expressions when the earth-resistivity varies exponentially with depth.

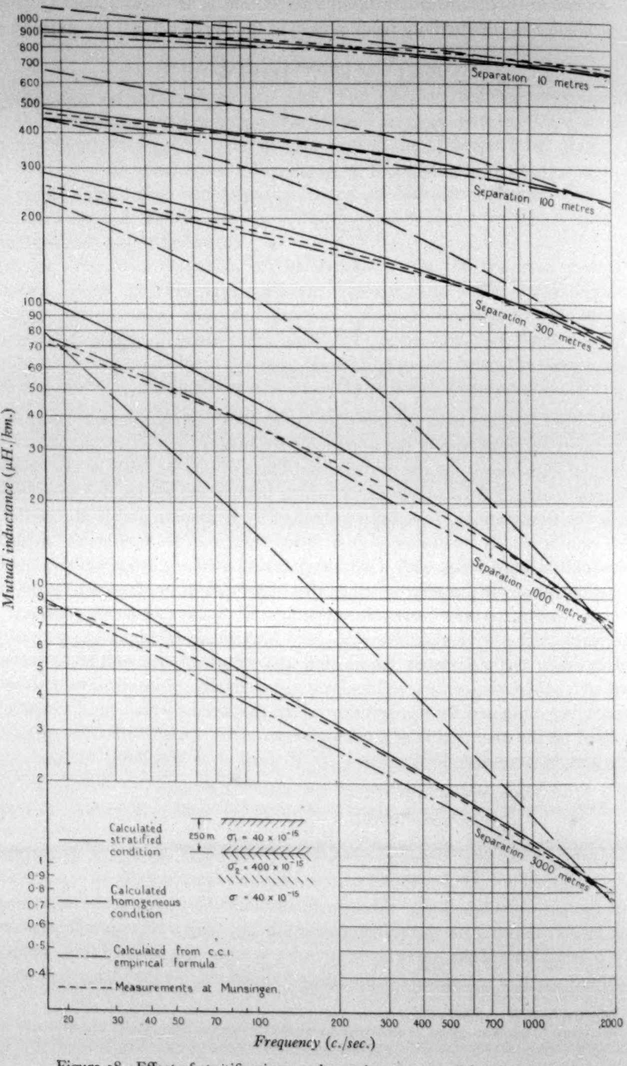


Figure 18. Effect of stratification on alternating currents (Munsingen tests).

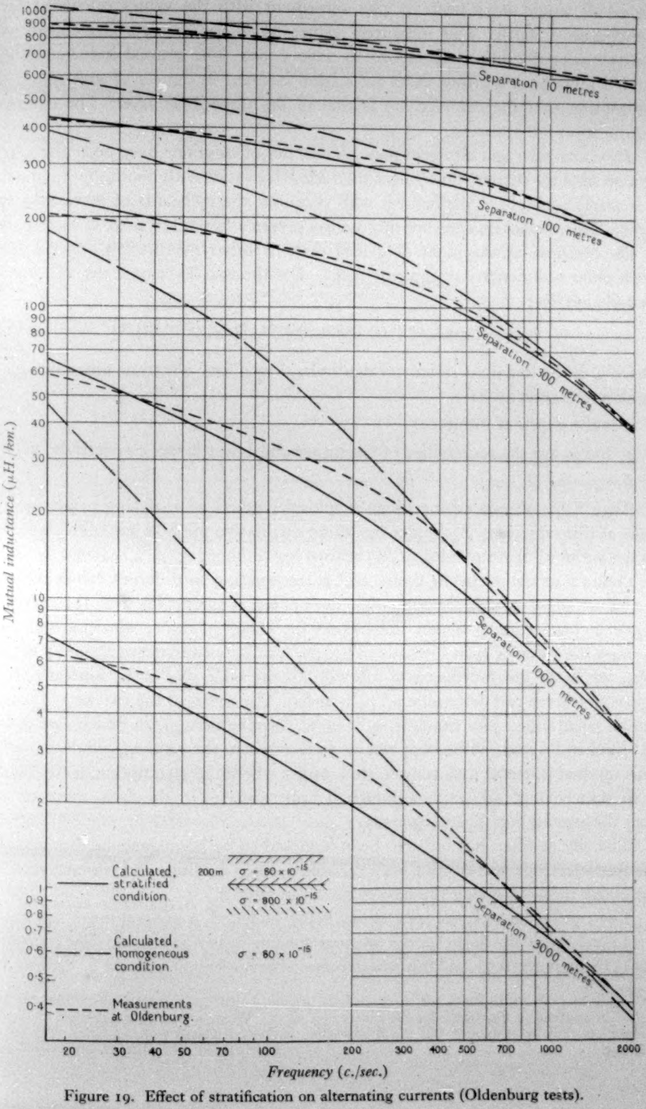


Figure 19. Effect of stratification on alternating currents (Oldenburg tests).

the result would have been to give agreement with the values calculated for a homogeneous earth. The measured values gave approximate agreement with homogeneous conditions from 800 to 2000 c./sec. The mutual inductances calculated for a homogeneous earth have been inserted in figures 18 and 19, on the assumption that the conductivity is that of the uppermost layer. The effects of stratification are apparent.

Direct currents and divergent flow. A number of solutions have been given from time to time for current-flow where the resistivity varies with depth, but in practice it is rarely possible to distinguish with certainty the constants of more than two layers. This, however, does not impose too severe a limitation since it corresponds to the common arrangement of glacial drift or other overburden superimposed upon older sedimentary or igneous rocks. For the two-layer case the well-known formula applies:

$$\rho_a/\rho_1 = 1 + 4\sum_{1}^{\infty} k^n \{[1+(2n\,h/a)^2]^{-\frac{1}{2}} - [4+(2n\,h/a)^2]^{-\frac{1}{2}}\} \qquad \ldots\ldots(11),$$

ρ_a where ρ_a is the apparent resistivity determined by a four-electrode measurement at
a an electrode interval a;

h h is the depth of interface;

ρ_1 and ρ_2 are the resistivities of the upper and lower layers respectively, and $k = (\rho_2 - \rho_1)/(\rho_2 + \rho_1)$.

Tagg⁽⁹⁾ has shown an ingenious graphical method of analysing results on the basis of this equation. A simpler but more qualitative method has been employed by the authors. If we plot log (ρ_a/ρ_2) against log (a/h) or log (ρ_1/ρ_a) against log (h/a), we obtain a series of curves, figure 20,* corresponding to different values of ρ_2/ρ_1. If now we superpose an experimental curve of log ρ_a against log a on the curves of figure 20 (*a*) so as to obtain the closest fit, then the unity abscissa will pass through the value of ρ_2 in the scale of ρ_a, the number of the curve chosen will give ρ_1/ρ_2, while the unity ordinate of a/h will cut the scale of a at h. Similarly, with figure 20 (*b*), we can determine ρ_1, ρ_2/ρ_1 and h. The mean of the two sets of values can be taken unless, for example, ρ_1 is more variable than ρ_2, in which case ρ_2 can be found and a mean value of ρ_1 can be deduced from the ratio ρ_1/ρ_2 and *vice versa*. This method is useful with complicated results where the assumption of two layers is to some extent arbitrary, although it approximates to the facts more closely than the assumption of homogeneity.

The site at Shap has hitherto been assumed to be homogeneous, but actually the four-electrode survey gave very variable surface readings and was analysed by the above method, giving the results shown in table 4.

The results for unequal separations being included, a general mean amounted to 250,000, while the mean for the unequal-spacing tests was 210,000 (low resistance tracks being neglected), and this was in reasonable agreement with the a.-c. measurements.

---

* The method is described for the case when $\rho_2 > \rho_1$. When $\rho_1 > \rho_2$, K is negative. A similar method can still be employed but the fundamental curves calculated from equation (11) will be different.

Table 4

Test No.	Surface soil (Ω.-cm.)	Intermediate layer (Ω.-cm.)	Lower layer (Ω.-cm.)	Depth of lower transition surface (metres)
1 and 2	35,000	35,000	250,000	30
3	85,000	85,000	—	—
5	100,000	45,000	400,000	20
7	250,000	150,000	300,000	45
8	110,000	160,000	250,000	40
9	50,000	50,000	150,000	40

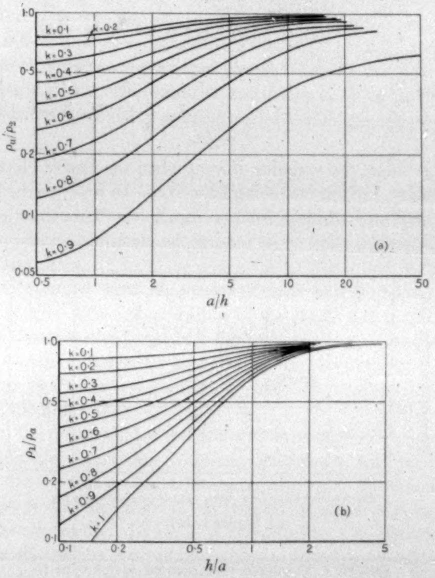


Figure 20. Variation of apparent resistivity with depth for two layers.

Variation of resistivity in a horizontal direction is not so easily dealt with by the four-electrode method. A simple case is that of an earth of resistivity ρ_1 separated by a vertical surface of separation or fault from an earth of resistivity ρ_2. If $\rho_1 = \lambda\rho_2$, the line of electrodes is normal to the fault, ρ_a is the apparent resistivity for an electrode-interval a, and l_1 is the distance of the fault from the first current electrode situated in the direction of the medium that has the resistivity ρ_2, then the following results can be proved:

λ, ρ_a

a

(i) All electrodes in medium 1.

$$\frac{\rho_a}{\rho_1}=1+\frac{\lambda-1}{\lambda+1}\left\{\frac{1}{2x+1}+\frac{1}{2x+5}-\frac{1}{2(x+1)}-\frac{1}{2(x+2)}\right\} \quad \text{......(12.1).}$$

(ii) Boundary between first two electrodes.

$$\frac{\rho_a}{\rho_1}=\frac{\lambda+3}{2(\lambda+1)}+\frac{\lambda-1}{\lambda+1}\left\{\frac{1}{2(2-x)}-\frac{1}{5-2x}\right\} \quad \text{......(12.2).}$$

(iii) Boundary between two potential electrodes.

$$\frac{\rho_a}{\rho_1}=\frac{\lambda^2+1}{\lambda(\lambda+1)}+\frac{\lambda-1}{\lambda+1}\left\{\frac{1}{\lambda(2x-1)}-\frac{1}{5-2x}\right\} \quad \text{......(12.3).}$$

(iv) Boundary between last two electrodes.

$$\frac{\rho_a}{\rho_1}=\frac{3\lambda+1}{2\lambda(\lambda+1)}+\frac{\lambda-1}{\lambda(\lambda+1)}\left\{\frac{1}{2x-1}-\frac{1}{2(x-1)}\right\} \quad \text{......(12.4).}$$

(v) All electrodes in medium 2.

$$\frac{\rho_a}{\rho_1}=\frac{1}{\lambda}+\frac{\lambda-1}{\lambda(\lambda+1)}\left\{\frac{1}{2x-1}+\frac{1}{2x-5}-\frac{1}{2(x-1)}-\frac{1}{2(x-2)}\right\} \quad \text{......(12.5).}$$

$$x=l_1/a.$$

Figure 21 (*a*) shows the variation if $\lambda=4$ when the line of electrodes is moved across the boundary, the interval being kept fixed. In practice the boundary could not be determined to within a distance much less than the electrode-interval. Figure 21 (*b*) shows the effect of increasing the electrode-separation in the neigh-

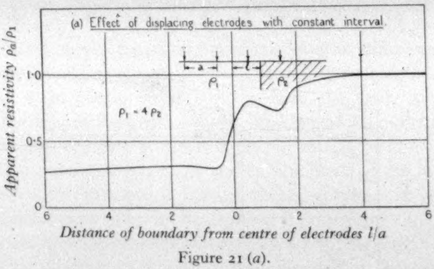


Figure 21 (*a*).

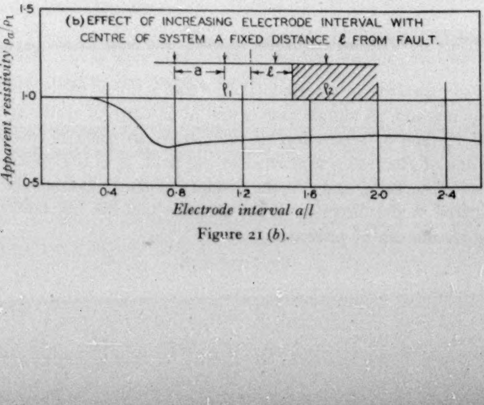


Figure 21 (*b*).

bourhood of such a discontinuity. We observe that its effect is very much less than the difference in the resistivities would suggest. For this reason the method of variable separation, although valuable for exploring horizontal stratification, is not satisfactory as regards vertical boundaries. If two fixed current electrodes are employed and if the equipotential lines are explored, then such a discontinuity can be located with relative ease and accuracy, as has been shown by Schlumberger(10).

The complexity of most geological formations has led to the use of empirical rules rather than analysis. The best known is that which states that a horizontal surface of discontinuity occurs at a depth corresponding to that separation for which the curve showing apparent resistivity against electrode-separation has a maximum curvature. In the simple case of two layers, which can be tested, this rule is not valid, but in more complicated cases it is useful in making rough estimates. In the authors' work, which was mainly concerned with the prediction of the linear flow of alternating currents from surveys made with d.-c. or divergent flow, the arbitrary assumption of two layers generally gave a sufficient approximation.

§4. GEOLOGICAL STRUCTURE

Although the method of sampling from borings and testing outcrops yields little information of value, the resistivity of a given rock formation in situ has a fairly constant value irrespective of geographical situation and, to a large extent, of depth. Furthermore, a wide range of values is observed extending roughly from 100 to 10^6 Ω.-cm. The determination of the shape and nature of geological formations by electrical methods requires methods of a higher accuracy than those described here. On the other hand prospecting, which is mainly concerned with the location rather than the recognition of bodies by absolute values, does not offer such great difficulties.

The converse, the estimation of resistivity from geological structure, is usually easier. Problems in connexion with telephone interference, with which the authors have been concerned, require only a low accuracy, since the mutual impedance between two lines is only sensitive to the resistivity where it is small, owing to separation and low resistivity. In addition the effective a.-c. resistivity for linear flow is much less affected by local irregularities and surface peculiarities than for d.-c. or divergent flow. Accordingly the Electrical Research Association with the assistance of Mr Broughton-Edge have prepared a map of Great Britain showing the resistivity corresponding to about 500 ft. in depth subdivided into ranges of (i) less than 1000 Ω.-cm., (ii) 1000 to 3000 Ω.-cm., (iii) 3000 to 30,000 Ω.-cm., (iv) 30,000 to 300,000 Ω.-cm., (v) greater than 300,000 Ω.-cm. This map, now published in two sheets, is based mainly on geological structure and measurements made by Mr Broughton-Edge and the authors, and on the results obtained by other workers.

§ 5. CONCLUSIONS

(*a*) Field experiments have shown that when the geological structure approximately furnishes the conditions for uniform resistivity, the distribution of alternating currents in the earth is that given by theory in the region of parallel flow. This being so, the average value of the earth resistivity over a large tract may be determined from measurements of the induction. Either the longitudinal e.m.f. induced in a second parallel line or the transverse e.m.f. induced in a search coil may be used.

(*b*) Divergent-flow methods, such as the four-electrode method, will give results agreeing with (*a*) subject to a rather wide margin on account of topographical variation leading to a dispersion of results. Skin effect must be avoided if a.-c. is employed, while polarization of the electrodes and ground causes corresponding limitations with d.-c.

(*c*) Within limits the resistivity can be related to the geological formation, and where the latter is uniform the resistivity has been shown to be independent of frequency up to 3000 c./sec. at least. The anomalous variation of apparent resistivity with frequency observed in earlier continental tests has been shown to be due to horizontal stratification.

(*d*) Stratification, at least as regards two horizontal layers, can be satisfactorily investigated by the methods of either (*a*) or (*b*).

§ 6. ACKNOWLEDGMENTS

The authors wish to acknowledge the assistance of the several investigators and organizations who assisted in the work described and particularly Mr J. E. Pidgeon (G.P.O.), Dr Klewe (Deutsches Reichspostzentralamt), who co-operated in connexion with the Reichspost testing-vans at Shap and Eltham, the late Mr Morgan and Mr Taylor of the Electrical Research Association, Dr J. Collard, whose results are quoted in the text, and Mr A. Broughton-Edge, who prepared the resistivity map.

REFERENCES

(1) Whitehead, S. *Phil. Mag.* **11**, 897 (1931).
(2) Pollaczek, F. *Elekt. Nachr.-Tech.* **3**, 339 (1926).
(3) Carson, J. R. *Bell Syst. tech. J.* **5**, 539 (1926).
(4) Foster, R. M. *Bell Syst. tech. J.* **10**, 264 (1933).
(5) Wenner, F. *Bull. U.S. Bur. Stand.* **12**, 469 (1916).
(6) Klewe, H. *Elekt. Nachr.-Tech.* **6**, 467 (1929).
(7) Josephs, H. J. *P.O. elect. Engrs' J.* **27**, 61 (1934).
(8) Riordan, J. and Sunde, E. R. *Bell Syst. tech. J.* **12**, 162 (1933). Also Gray, Marion C. *Physics*, **4**, No. 2, p. 76 (Feb. 1933).
(9) Tagg, G. F. *Min. Mag.*, Lond., **43**, 305 (1930).
(10) Schlumberger. *Etude sur la Prospection Electrique du Sous-sol.* Gauthier-Villars et Cie, Paris (1920).
(11) Radley and Whitehead. *J. Instn elect. Engrs*, **74**, 201 (March 1934).

537.322.1

THE DIRECT MEASUREMENT OF THE PELTIER COEFFICIENT

By A. J. WOODALL, B.Sc., Ph.D., A.Inst.P., Queen Mary College

Received April 4, 1935. *Read in title May* 3, 1935.

ABSTRACT. The fundamental problems in the direct measurement of the Peltier coefficient are discussed with reference to the various methods available. A new apparatus is described which includes all-metal calorimeters, separate from the specimens, and a differential platinum resistance thermometer. With a sufficiently good thermostat this would just detect a steady heating of the order of 10^{-7} cal./sec. The high sensitivity of this arrangement has permitted a strict test of the proportionality of the Peltier effect to current to be carried down to 0·012 A., with copper-constantan junctions. The error due to thermal losses along the leads has been especially studied, and estimates of the loss have been made.

---

§ 1. INTRODUCTION

Interest in the thermoelectric effects has been revived by recent developments in the electronic theory of metals; but the technique of measurement of the Peltier effect, in particular, has been very little advanced for nearly two decades. Moreover, no adequate critical account of the problem appears to exist. It therefore seemed advisable to include in this study of the problem a review of previous methods, not for historical interest so much as to help to direct attention to the variety of methods available, and more especially to the faults to which they are subject.

The bulk of this paper, however, consists of the description of a new method of measurement developed by the author.

§ 2. FUNDAMENTAL PROBLEMS

Before considering any methods in detail, it will be useful to glance at the difficulties which are of most general occurrence.

Let us suppose that the apparatus used has each junction to be studied enclosed in some form of calorimeter. This will also contain a length of each of the specimens joined, and may also have a heating-coil and some thermometric device.

A single measurement with such an arrangement will not discriminate between the contributions made by the Peltier effect and those due to any other thermal effects arising inside the calorimeter. The following are the characteristics of the various phenomena concerned.

Ideally, the Peltier effect is localized, and is reversible with current. In practice, however, in making a junction, transition layers are created through which the effect is more or less diffused.

The Joule effect is irreversible and is diffused.

The Thomson effect is diffused: but while it is reversible with current when the temperature-gradient remains the same, it also reverses with the latter. Hence it will make an irreversible contribution when the Peltier effect is the most important within the calorimeter. In this case it will be indistinguishable from Joule effect.

Finally we must remember that there will be a greater loss of heat along a given lead when the effect concerned takes place in that lead than when it does not. In particular, the apparatus will not be equally sensitive to heat supplied by the heating-coil and to that developed at the junction. In this connexion also, the presence of transition layers causes some of the evolution or absorption of Peltier heat to occur nearer the ends of the calorimeter, so that a larger fraction will be conducted along the specimens.

These points should be borne in mind when considering any particular design.

§ 3. REVIEW OF METHODS

There are various methods which are of historical importance but are either only qualitative, or not capable of any high accuracy. We may quote Peltier's cross, and the use of a differential air thermometer.

Le Roux(1) used mercury thermometers in water calorimeters. His results are still sometimes quoted, although much more accurate methods have been devised since.

Jahn(2) used the Bunsen ice calorimeter, studying one junction at a time. The Peltier effect was disentangled from the Joule heating by reversing the current, the latter effect being in fact much the larger. This is in itself a disadvantage and diminishes accuracy, while the limitation to one temperature is serious.

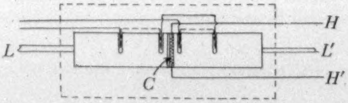


Figure 1. Pellat's apparatus.

Figure 1 illustrates apparatus described by Pellat in 1901(3), but apparently never used. It is indeed not very practicable as shown, but it introduced some new features which have reappeared in later work, and are worthy of discussion. Stout leads L, L' of the materials to be studied are joined to blocks of the same substances of large cross-section, the whole being in an evacuated enclosure to reduce heat-losses. Embedded in one block near the junction is a heating-coil C (leads at H, H'); while a set of thermojunctions is disposed in some such manner as that shown. The current in the heating-coil is to be adjusted so as just to compensate the Peltier cooling at the junction.

Since there will be a general Joule heating in each specimen, the thermojunctions are connected to give the difference between two mean temperatures: the average at the centres of the blocks, and that in the immediate neighbourhood of the junction. Pellat was prepared in addition to have cross-sections proportional to resistivities in an effort to eliminate more completely the effects of Joule heating. Even this precaution would probably prove inadequate with the currents such apparatus would require before measurable temperature-gradients could be established. It would also take a long while to reach the steady state.

There is yet another defect in this and similar methods. Thermoelectric properties of a material are very susceptible to changes in the mode of preparation and treatment, and there will in general be an appreciable Peltier coefficient between a wire and a block of the same metal. Thus Pellat's arrangement has virtually three junctions making unequal contributions to the recorded result. This leaves considerable doubt as to how far the value is representative of block metal, and how far of wire.

Cermak(4) made use of a thermoelectric calorimeter, figure 2, due to Lecher(5), in which a ring of thermoelements surrounded the junction (*J*) and served to indicate the temperature-difference between the liquid *A* and the reference-plane *B*. The heating-coil *K* was used to calibrate the arrangement. The leads *H*, *L* were of the one material and *C* of the other. A modification having liquid metals in hard glass tubes enabled Cermak to investigate the change in the Peltier coefficient ensuing on change of state of one component. The results obtained were of considerable value, especially as the range of temperature was very extensive. The measurements were not of the greatest precision, as shown for example by the variations in successive values at the same temperature. This method, like most others, is liable to serious systematic error arising from preferential thermal conduction along the leads.

Another method using calorimeters containing liquid was introduced by Barker(6) and by Caswell(7), the two junctions being each immersed in a Dewar vessel containing also a stirrer, a thermoelement, and a heating-coil. The coil was used on the side of Peltier cooling, the adjustment aiming at a net evolution of heat equal to that arising in the other calorimeter. This requires a supply of energy double that involved in the Peltier effect.

The work was carried out necessarily in a restricted space and presented some difficulty in consequence, although it was hoped to obtain great sensitivity by the use of Dewar flasks. Stirring is of very great importance, and irregularities in stirring are very serious. For example, Caswell reported(8) that one recorded value fell to one-half when stirring ceased, while irregularity seriously upset the steady temperature-distribution. Depth of immersion was another factor which affected the results, while evaporation proved troublesome when mobile oils were used.

Caswell also tried another method, in which resistance coils were wound in series with each junction, and one was shunted so as to carry less current. The extra Joule heating at the other junction was to compensate the Peltier cooling there. The method was abandoned on account of stirring and other troubles, not because of anything inherently wrong with it.

Beck(9) introduced an indirect method in which the steady temperature-distribution in a system of rods carrying a current was analysed to give the Peltier coefficient, the thermal conductivities being assumed. Later workers seem to have avoided the method, which is difficult and tedious, apart from its dependence on other experiments for some of its data.

Figure 3 represents one half of a symmetrical apparatus used by Jordan(10) arranged for measurements for copper against bismuth; the block *B* and the lead *L* were of copper, the rod *R* of bismuth. It was found convenient to make the heating-coil *C* share one lead (*L*) with the junctions. Temperature-differences between the pair of blocks were indicated by thermojunctions set in grooves *Th*. Four thermocouples were used in series to give greater sensitivity. The principle of the method of measurement is the same as that described above for Barker and Caswell. This is a compact and sensitive form of apparatus, and was actually designed for use between pole-pieces. The actual temperature of each block was given by a further thermojunction (not shown).

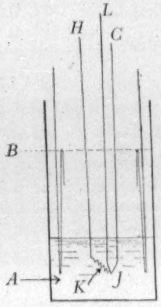


Figure 2. Part of Cermak's apparatus.

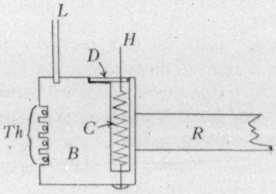


Figure 3. One side of Jordan's apparatus.

This method does not differ essentially from that used later by Gottstein(11). This worker treated the heat-flow at the ends of the rod (corresponding to *R* above) theoretically. However, the existence of an unknown Peltier effect at the junction of the lead, and the awkward shape of a block recessed for a heating-coil, render it almost impossible to deal adequately with the flow throughout the block.

It is very necessary to consider carefully the heat-losses of any arrangement, and the disc *D*, figure 3, was introduced by Jordan to diminish the loss of heat along the lead *H*. Gottstein appears to have been mainly interested in the loss along *R*.

Finally the convenient and rapid relative method of Borelius(12) must be mentioned. This uses blocks with a wire of the specimen *X* spanned between them.

In terms of the Peltier coefficient between reference metals A and B, the measurements yield the coefficients of X against both A and B, and the thermal conductivity of X.

This selection of methods, while not exhaustive, is representative of the methods in use up to the commencement of this research. Quite recent work on single crystals, for example, has been carried out by the Barker-Caswell method(13).

§4. SYSTEMATIC ERRORS

Probably the most serious systematic error in most methods is due to the preferential thermal conduction mentioned in §2. The usual methods are valid only if the body (calorimeter) whose temperature is studied receives a constant fraction of any heat developed within it whatever the source. For example, in many methods the leads to the junction are comparatively stout and are good conductors; and there will usually be a much greater tendency for heat to pass along them to or from the junction than when the energy is dissipated in a heating-coil more or less remote from them. This will affect especially the Cermak and Barker-Caswell methods, and others with similar arrangements. Most frequently the preferential transfer will be greater along the junction-leads than along the heating-coil leads, and the measured Peltier effect will then be too low.

Where intermediate metal is used a further complication is introduced by the multiplication of junctions and hence of Peltier effects. A fairly complete knowledge of the heat-flow is then necessary in order to predict the connexion between the various Peltier effects and the temperature or temperatures of the selected point or points of the calorimeter. The Pellat, Jordan, Gottstein and Borelius arrangements can all be criticized on this score. It is important to realize also the advantage of measuring average temperatures rather than temperatures at a point when a complete account of heat-flow cannot be given. Thus Jordan's use of several thermojunctions scores on this account as well as by virtue of the increased sensitivity. A platinum resistance thermometer would be a still further advance.

It will be realized that it is extremely difficult to estimate the error arising in the Jordan-Gottstein type of apparatus from preferential thermal conduction. In methods in which liquid calorimeters are used, it is somewhat easier to investigate the matter, at least qualitatively, by making arrangements to change the depths of immersion of the various sources of energy. In general less heat will escape preferentially when the appropriate heating-source is immersed more deeply.

The liquid calorimeter provides another method of varying the proportion of heat escaping along the leads, namely by altering the efficiency of stirring. In point of fact the stirring provides one of the main difficulties of the method, as violent agitation is often awkward to arrange and to maintain constant, and the energy involved may easily exceed the Peltier effect to be studied.

The method described in the next section deals with most of these points; and although preferential thermal conduction is not entirely eliminated, it has been found possible to estimate it approximately, and to control it to some extent by modifications of the apparatus. No intermediate metal is involved, except that due to soldering or welding, and the specimens remain straight and undeformed. There is no liquid in the calorimeter; and an average temperature is recorded by a platinum resistance thermometer.

§ 5. THE AUTHOR'S METHOD

The general arrangement may be understood by reference to figure 4, which is a horizontal section through the thermostat used. This was of copper (*B*), lagged, and surrounded by a wooden box *W*. Close parallel copper tubes *TT′* ran horizontally right through it, being jacketed by a cylindrical water bath *J*. The calorimeters were placed centrally in these tubes at *KK′*, and the various leads passed out in bundles *LL′*. *T″* represents the bulb of a thermometer giving the temperature of the jacket at a point between the tubes. The other symbols are explained beneath the figure. The thermostat arrangement is not ideal, and this did in fact severely hamper the work; but it was good enough to enable the merits of the method to be established.

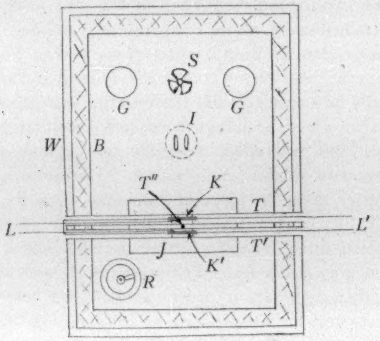


Figure 4. Plan view of thermostat. *S*, stirrer; *GG*, regulating lamps; *I*, immersion heater for rapid heating; *R*, spiral aniline-mercury regulator.

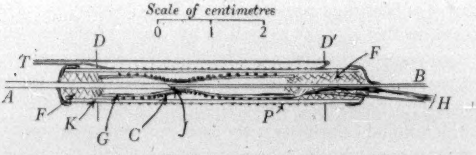


Figure 5. Calorimeter.

The latest form of calorimeter used is represented in figure 5. A copper tube *K* (length 5·5 cm., external diameter 6 mm., thickness of walls 0·6 mm.) has the specimens *A*, *B* lying along its axis, with the junction at *J*. A glass tube *G*, drawn down very thin and shaped so as to fit the wires snugly, carries the manganin heating-coil *C*. Current and potential leads for this coil pass out at *H*.

The calorimeter is supported coaxially in the copper tubes (*TT′* of figure 4) by stiffened paper discs *DD′*, and between these is wound the platinum resistance coil *P*. This is of fine silk-covered wire (s.w.g. 44, resistance 20 Ω. per metre), insulated by a layer of pure well-baked shellac from the copper calorimeter *K*. The leads are shown at *T*. *FF* is packing of high thermal conductivity, consisting of tin-foil and copper spirals insulated with shellac.

A reliable thermoelectric potentiometer (made by the Cambridge Instrument Co.) formed the basis of the electrical measurements, in conjunction with a galvanometer of sensitivity quite adequate for the purpose, since it gave a deflection of 12·8 cm. per microampere at about $2\frac{1}{4}$ metres. Its resistance was 12·8 ohms. Potential differences up to 90 mV. could be estimated to about 1 μV., the limit of reading of the dial. The makers' certificate claimed that the instrument was not subject to errors greater than this. A Weston cell, checked against a standard cell, was used for this potentiometer.

Currents were determined by measuring the potential-differences across two standard coils, a one-ohm and a tenth-ohm coil respectively. The platinum resistance coils were used differentially, being adjacent arms of a Wheatstone bridge. The same galvanometer was used in this bridge as for the potentiometer; a system of keys enabled currents and potential-differences to be measured during the course of an experiment. Preliminary readings of currents were made on milliammeters to avoid giving the galvanometer violent kicks in the middle of readings, with consequent shift of zero.

It was also possible to switch the two platinum coils, in series, into another bridge for studying their mean temperature. Recorded temperatures were, however, always referred to a sub-standard mercury-in-glass thermometer which had been checked against an instrument with National Physical Laboratory certificate. The accuracy of temperature-measurement was actually somewhat greater than was essential.

The method of making a determination is broadly as follows. The thermostat is set to the required temperature, and left to settle down. Meanwhile the currents are adjusted approximately, if a rough value of the coefficient is known. When the galvanometer spot is steady, indicating that the fluctuations in temperature of the bath and other disturbances have died down, readings of the galvanometer are made against time. The current through the junctions and the current through the appropriate coil are started simultaneously, and rough measurements of them are made in between the deflection-time readings. The Peltier heating and cooling take a little longer to affect the calorimeter than the heating-coil, and consequently the initial movement of the spot is followed by a reversal and a final settling down to a steady value after a period of 10 to 20 minutes. The final readings of current are then made. The currents are next switched off, and the calorimeters allowed to regain their initial temperatures. The experiment is now repeated with the current in the junctions reversed and the other heating-coil in use. If time presses, this may follow the first part at once, but the intermediate reading of the zero position is useful.

A second experiment is then made with one current altered. If possible the deflection should be of the same order as before, but in the opposite sense; both deflections being small. Interpolation then gives the current for a perfect balance. This corresponds to equal net heating on the two sides, and, as in similar methods, the Peltier effect is equal to one-half the energy supplied by the coil.

In practice there are several variants which may be made in the procedure. For example, a rough adjustment of the currents may be made during the course of an experiment. Again, it is usually advisable to obtain figures for the (steady) deflections with given currents in one coil and with no junction-current, as these are useful in estimating the change in current necessary to produce a given change in deflection.

Earlier arrangements were somewhat different from that described, and a comparison of results is of some value as the progressive reduction of thermal conduction losses is clearly shown. The set of results in table 1, for the Peltier coefficient of a $\frac{1}{2}$-mm. pure copper wire against a 1-mm. wire of pure constantan at 20° C., refers to a series of different forms of calorimeter.

Table 1

Apparatus no.	1	2	3*a*	3*b*
Coefficient (volts)	0·00990	0·01007	0·01058	0·01100

Between 1 and 2 the modifications could not be expected to affect the conduction very greatly, although there should be some improvement; considerable reduction in the losses should have resulted from the change to 3, the form described above. 3*a* and 3*b* refer to successive improvements in the packing at the ends of the calorimeters. The figures quoted are entirely in agreement with what would be anticipated from considerations of design.

It was arranged in the later work that the thermal contact between the one (copper) wire to the junction, the leads to the heating-coil, and the walls of the calorimeter, should be as intimate as was practicable consistently with electrical insulation. Hence it was hoped to reduce thermal losses along these wires to very small values.

Now the residual loss presumably varies according to the local temperature conditions in the calorimeters. The design left little likelihood of any appreciable leakage along the heating-coil leads; so that we have to consider only the copper wire. (The other wire, of constantan, would be of relatively small importance.) Now when there is Peltier cooling, and compensation from the coil, it is very unlikely that the junction will gain appreciable heat along its lead, owing to the close proximity of a source of heat. However, when there is Peltier heating at the junction this no longer applies, and one would expect some loss.

In the same way, experiments on single junctions, with and without compensation by heating-coil, provide results differing by a factor representing the fraction of heat transferred along the copper wires on the two sides. Thus it proved possible, by a series of experiments, to obtain an estimate of the order of magnitude of this fraction, as well as separate values for sensitivity and for Joule and Peltier effects on the two sides of the apparatus. It was, for example, possible to detect differences in recorded Peltier effect due to projection from one calorimeter of a short length of wire heated during the soldering of the junction.

The fractions of heat lost, as given by this method, ranged about 2 per cent; but the true values are probably somewhat larger. According to this argument, the results of the usual null method should be too low by about half this fraction.

§ 6. NUMERICAL RESULTS

Peltier coefficients of the order of a 0·01 V., such as those of copper-bismuth and of copper-constantan, must be reckoned large. Hence most workers have found it necessary to make measurements with currents from about $\frac{1}{3}$ A. upwards. This means that the Joule heating effective in the apparatus is usually of the same order as the Peltier effect, and may even be considerably greater. It can be demonstrated that the usual repetition of the experiment with current reversed only eliminates error due to the irreversible heating if there is perfect symmetry. Moreover, very small Joule heating ensures fulfilment of the condition of § 2 for neglect of the Thomson effect.

These are all strong reasons for seeking a method of high sensitivity; and only with such a method is it possible to test the proportionality of Peltier effect to current over a wide range, or to measure small Peltier coefficients accurately.

The maximum sensitivity of the apparatus described has never been attained; but as it was last used, a heating or cooling of 10^{-4} watt on either side produced a steady deflection of 1·8 cm. This implies a deflection of $\frac{1}{20}$ mm.—the minimum detectable—for a heating of under 3 erg/sec. In other words, the Peltier coefficient for copper-constantan, which is just over 0·01 V. at ordinary temperatures, could be measured to an accuracy of 1 per cent with a current of the order of 1 mA.

The bridge can be readily modified to increase the sensitivity; but it must be appreciated that this work makes extraordinarily stringent demands on the thermostat. For example, to take full advantage of the sensitivity detailed above, the temperatures of the places selected for the calorimeters must remain equal during the course of an experiment to an accuracy of about 0·0001° C. This is well beyond the capacity of the thermostat used in this work.

Table 2 gives the values deduced for the Peltier coefficient, again for copper against constantan at about 20° C., for different currents. It is a test of propor-

Table 2. Peltier coefficient with different currents

Temperature (° C.)	20·19°	20·05°	20·00°	20·06°	20·10°	20·00°
Current (A.)	0·01168	0·0301	0·0301	0·0500	0·0500	0·1000
Peltier coefficient (V.)	0·01053	0·01059	0·01056	0·01058	0·01064	0·01059

tionality of the Peltier effect to current, and at the same time it may be regarded as an investigation of the observational accuracy of the method, since some of the determinations are repetitions of others with the same currents, made after intervals of a few days.

All these results are subject to thermal conduction losses amounting to the same fraction in each case; they are all therefore somewhat lower than the true coefficient at this temperature.

There is no evidence of any systematic variation of the coefficient with current in this series. Moreover, the results show that, although the observational accuracy did not reach the limits indicated above, it was high for the comparatively small currents used.

The best measurements were over a somewhat limited range of temperature, 17·5° to 38° C. The results are expressible in the form

$$\begin{aligned}\Pi &= 0{\cdot}01100 + 0{\cdot}000081\ (t-20)\\ &= 0{\cdot}01100\ \{1 + 0{\cdot}0073\ (t-20)\},\end{aligned}$$

in which Π is the Peltier coefficient in volts, and t the temperature in degrees centigrade.

The thermo-e.m.f. E of the couple in volts was found to obey the law

$$10^3 . E = 0{\cdot}03543 . t + 0{\cdot}000084 . t^2$$

over the range 25° to 58° C., so that

$$\begin{aligned}T\frac{dE}{dT} &= 0{\cdot}01136 + 0{\cdot}000084\ (t-20)\\ &= 0{\cdot}01136\ \{1 + 0{\cdot}0074\ (t-20)\},\end{aligned}$$

where T is the temperature in degrees Kelvin.

Thermal conduction would tend to depress the measured value of Π, but it should not affect the value of the temperature coefficient expressed as a fraction of the coefficient measured at some fixed temperature. It will be seen that the measured values of Π are about 3 per cent below those predicted by the relation $\Pi = T . dE/dT$, while the temperature coefficients, each divided by the corresponding value of Π at 20° C., agree almost exactly.

§ 7. DISCUSSION OF THE METHOD

We proceed to discuss briefly the advantages and limitations of this method.

First we may note that the specimens used can remain straight and undeformed; and that no other material is interposed, except solder. The calorimeter is designed to minimize the amount of heat lost along the leads, so that the most serious systematic error is reduced to a small quantity. It may also be remarked that some estimate of the residual error is possible. This work appears, in fact, to be almost the only research in which a serious attempt has been made to deal with this fundamental trouble.

The sensitivity of the apparatus is high, and it can be increased when necessary by a simple adjustment of the resistance bridge. It will be noted that the use of a platinum resistance thermometer is largely responsible both for this high sensitivity and for the ease with which it can be varied, and has the further advantage of making the measurements depend on average temperatures rather than on the temperatures of isolated points of each calorimeter, as thermocouples do.

The apparatus is at present adapted for wire specimens only; but at least it avoids the ambiguity of those methods which have both wire and block metal in their construction and yield results representative of neither. In modern work this should be a point worthy of consideration.

The method as described makes demands which are rather excessive for the usual form of thermostat. Both calorimeters could be used in one tube, in line with one another. This should improve the performance, although the multiplicity of leads passing along the tube will introduce difficulties.

In conclusion it must be pointed out that the dimensions of the calorimeter given in § 3 were not chosen as the best possible, but as the most convenient for studying the changes due to modifications in packing. For greater accuracy the tubes should be somewhat longer, and if practicable a little narrower. This will cause a slight loss in sensitivity which will, however, be more than offset by the reduction in systematic error.

§ 8. ACKNOWLEDGMENTS

I wish to record my gratitude to Prof. H. R. Robinson, F.R.S., who suggested the problem, and whose guidance has been valuable throughout; and to the Department of Scientific and Industrial Research and to the East London College (now Queen Mary College) for the grants and facilities which made the research possible.

REFERENCES

(1) Le Roux. *Ann. Chim. (Phys.)* **10**, 201 (1867).
(2) Jahn. *Ann. Phys.*, Lpz., **34**, 755 (1888).
(3) Pellat. *C.R. Acad. Sci.*, Paris, **133**, 921 (1901).
(4) Cermak. *Ann. Phys.*, Lpz., **24**, 351 (1907); **26**, 521 (1908).
(5) Lecher. *Ann. Phys.*, Lpz., **19**, 853 (1906); **20**, 480 (1906).
(6) Barker. *Phys. Rev.* **31**, 321 (1910); **34**, 224 (1912).
(7) Caswell. *Phys. Rev.* **33**, 379 (1911); **7**, 269 (1916); **12**, 226 (1918).
(8) Caswell. *Phys. Rev.* **33**, 391 (1911).
(9) Beck. *Vjschr. naturf. Ges. Zürich*, **55**, 103, 470 (1910).
(10) Jordan. *Phil. Mag.* **21**, 454 (1911).
(11) Gottstein. *Ann. Phys.*, Lpz., **43**, 1079 (1914).
(12) Borelius and Lindh. *Ann. Phys.*, Lpz., **51**, 607 (1916).
Borelius. *Ann. Phys.*, Lpz., **52**, 398 (1917); **53**, 615 (1917).
(13) Fagan and Collins. *Phys. Rev.* **35**, 421 (1930).

537·531·8: 535·33·07

COMPARISON OF X-RAY WAVE-LENGTHS BY THE PLANE-GRATING VACUUM SPECTROGRAPH, AND THE STRUCTURE OF THE K LINE OF CARBON

By D. P. R. PETRIE, M.Sc., Research Scholar,
University of Melbourne

Communicated by Prof. T. H. Laby, February 12, 1935. *Read in title May* 17, 1935

ABSTRACT. The relative merits of plane- and concave-grating spectrographs for precise measurement of wave-length between 10 and 100 A. are discussed. It is shown that experimental conditions do not allow full advantage to be taken of the superior resolving-power of the concave grating. A plane-grating spectrograph was designed for relative wave-length measurements of high precision; its dispersion and resolving-power are calculated, and a number of wave-length determinations relative to Cu $L\alpha_{12}$, viz. 13·310 A., are reported. Microphotometer curves of the K line of carbon show that this line has, besides the principal line, two components of shorter wave-length. This result is in good accordance with that of other workers.

§ 1. INTRODUCTION

The precise absolute determination of wave-lengths in the region of the so-called "ultra-soft" X-rays (say 10 to 100 A.) has claimed the attention of many experimenters during the last decade, and a satisfactory solution of the problem has yet to be made. It is hardly necessary to stress the need for such measurements to provide accurate data for the further development of theoretical spectroscopy. In addition, however, Laby and Bingham⁽¹⁾ have pointed out that, by determining absolutely wave-lengths which have also been measured by the
d crystal method, the lattice constant *d* of the crystal used may be calculated from the equation

$$n\lambda = 2d \sin \theta.$$

Since *d* is known in terms of the Faraday, the electronic charge, and the density and molecular weight of the crystal, such a procedure would lead to an evaluation
e of the electronic charge *e*, which is the least accurately known quantity in the expression for *d*.

The chief difficulty in carrying out such work is in obtaining a suitable spectrometer. The use of an interferometer for these rays being apparently inadmissible, the choice lies between the plane-grating and the concave-grating spectrographs. Although the concave grating is usually preferred on account of its theoretically higher resolving-power, dispersion and intensity, there are serious drawbacks to its use in this region of the spectrum.

Some consideration will now be given to the relative merits and defects of the two instruments. The resolving-power of a concave grating used at grazing angles of incidence has been discussed theoretically by Mack, Stehn and Edlen⁽²⁾, while that of a plane grating used at grazing incidence with divergent light has been given approximately by Prins⁽³⁾. According to Prins, if y is the distance of any grating-element from the central one O, the path-difference between a ray striking O and one striking P is expressible as a power series in y. If parallel light is used with a plane grating, the quadratic and higher-power terms vanish, and there is no theoretical limit to the resolving-power. If, however, the light diverges from a point source, as in the case of all X-ray spectrographs, Prins states, as a criterion for best resolution, that y should not exceed the value Y_{opt} which makes the quadratic term equal to half a wave-length. Thus the length of ruling which can be used, and hence the resolving-power, is seriously limited.

In the case of the concave grating, Mack, Stehn and Edlen have shown that, owing to the focusing conditions, the second- and third-power terms vanish. By equating the fourth-power term to half a wave-length a much larger value of Y_{opt} is obtained, on account of the relative smallness of this term. Hence a concave grating of optimum width possesses a greater resolving-power than a plane grating of optimum width, the source being assumed to be infinitely narrow in both cases.

Mack, Stehn and Edlen, however, point out that the resolving-power actually obtained in a spectrograph is, for wave-lengths below about 100 A., limited by the finite width of the slit. They show that the resolving-power of a slit varies proportionally to the wave-length, and in inverse proportion to the slit-width. For a wave-length of 5 A. and a slit-width of 0·01 mm., the resolving-power is only 27.

Table 1 gives numerical values for the following quantities: (*a*) The resolving-power in the first order of a concave grating of optimum width, used in conjunction with an infinitely narrow slit. The radius of curvature of the grating is 1 metre, there are 571 lines per mm., and the angle of incidence is 5°. (*b*) The resolving-power in the first order of a plane grating of optimum width, used in conjunction with an infinitely narrow slit. The source and image are each 50 cm. from the grating, there are 571 lines per mm., and the angle of incidence is 1°. (*c*) The slit-width which would have the same resolving-power as the plane grating.

Table 1

λ (A.)	*a*	*b*	*c* (μ.)
5	3100	530	0·5
20	4600	670	1·6
100	6900	740	7·3

The second and third columns show that the resolving-power of a concave grating *alone* is, under average experimental conditions, about 6 or 9 times that of a plane grating. However the fourth column indicates that, to render a concave-grating spectrograph superior to a plane-grating spectrograph in resolving-power, it would be necessary to use slit-widths which are prohibitively small in practice on account of clogging etc.

From the point of view of spectral intensity, it is at once evident that the concave grating is superior, for the intensity is proportional to the solid angle of rays which the grating receives from the source, and this solid angle is, for a given angle of incidence, proportional to the aperture of the grating; it has already been shown that the permissible aperture is larger in the case of a concave grating than in the case of a plane grating, the ratio being the same as the ratio of their resolving-powers, which in the above numerical example varies from 6:1 to 9:1.

The dispersion obtainable from the two instruments considered above is quite adequate to take advantage of all the available resolving-power. This is evident from table 2, in which the dispersion is expressed by the ratio $\lambda/\delta\lambda$, where $\delta\lambda$ is the wave-length interval corresponding to a distance of 0·004 mm. on the photographic plate, and is calculated from the well-known dispersion formulae. Thus $\lambda/\delta\lambda$ is a measure of the accuracy obtainable for a wave-length λ, on the basis that the position of the line on the plate can be measured to 0·004 mm. Although in the case of the concave grating the dispersion falls short of the theoretical resolving-power at lower wave-lengths, yet it is adequate in view of the limitation on resolving-power imposed by the slit-width.

Table 2

Wave-length (A.)	Concave grating	Plane grating
5	800	1200
20	2900	2800
100	10300	6600

The concave-grating spectrograph has a serious drawback in that diffracted rays meet the photographic plate at small glancing angles. Eddy and Laby, in their experiments on the concave-grating spectrograph, have found that, even for an angle of incidence of 5°, surface irregularities in the photographic plate and in the emulsion and the anticlastic curvature of the plate, which depends on the plate's thickness, seriously detract from the perfection of the lines on the plate, and set a limit to the smallness of the angle of incidence which may be used with the concave-grating spectrograph.

From the above considerations it may be concluded that the plane grating is the most convenient apparatus wherewith to investigate wave-lengths in the region under consideration, for the advantage of extra resolution afforded by the concave grating is outweighed chiefly by the difficulty of using sufficiently fine slits and sufficiently flat photographic plates.

§ 2. EXPERIMENTAL METHOD

Principle of spectrometer. The X-ray beam incident on the grating is not, as customarily it is, a narrow parallel beam collimated by two or more slits, but a wide beam diverging from a line source, placed near the grating, with a single slit placed close to the grating but on the side remote from the source.

The fine line source, which is itself equivalent to a second slit, is obtained by turning the target face of the X-ray tube until it subtends only a very small angle at the axis of rotation of the grating. The advantages of this method over the usual one are (i) a much higher intensity of the radiation incident on the grating, permitting of shorter exposures; (ii) absence of the difficulty encountered in the two-slit grating spectrometer in bringing the focal spot on to the axis of the slits.

Theory of the single-slit grating spectrometer: dispersion. The dispersion of the instrument will first be considered. In figure 1, S_0 is the source, GH the grating, and S_1 the slit. Since the incident rays form a divergent bundle so also will the diffracted rays of any particular wave-length, in accordance with the ordinary grating equation $n\lambda = D(\cos i - \cos\theta)$. The slit selects from the diffracted rays a narrow divergent pencil which falls on the photographic film, producing a line. n, λ, D, i, θ

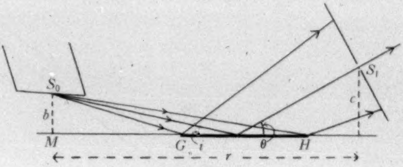


Figure 1.

The problem of finding the distribution of wave-lengths along the film is simplified by assuming that the film is constrained to lie on a circle whose centre is at the slit. Although this was actually the case in the earlier experiments in which films were used, the use of flat plates became necessary later. The same results will apply approximately to this case, however, as the curvature of the films was only slight.

In finding the relation between λ and θ we may assume the source and slit to be infinitely narrow, so that we need consider only infinitely narrow pencils. The rays of any particular wave-length passing through the slit are then regarded as being diffracted by an infinitely small section of the ruled surface. Then all rays reaching the film have to satisfy two conditions:

(i) The ordinary diffraction equation

$$n\lambda/D = \cos i - \cos\theta \qquad \text{......(1).}$$

(ii) They must pass through the two points S_0, S_1. This geometrical condition may be expressed thus:

$$b\cot i + c\cot\theta = r \qquad \text{......(2).}$$ b, c, r

Elimination of i gives the required relation between λ and θ:

$$n\lambda/D = [1 + \tan^2\theta/(R\tan\theta - a)^2]^{-\frac{1}{2}} - \cos\theta,$$

where $R = r/b$ and $a = c/b$. R, a

If θ is small, we have approximately

$$n\lambda/D = 1 - \cos\theta - \tfrac{1}{2}\,[\tan\theta/(R\tan\theta - a)]^2 \qquad \dots\dots(3).$$

Figure 2 shows the graph of this equation for typical values of the geometrical constants R and a, and also the graph of $1-\cos\theta$. It will be noted that the term $\frac{1}{2}\,[\tan\theta/(R\tan\theta - a)]^2$ is relatively small for large values of λ, but increases rapidly as λ approaches zero. Thus the dispersion (the slope of the curve) is never as great as, but often approximates to, that of the usual type of spectrograph, where two collimating slits are used. The dispersion of the latter is given by the slope of the $(1-\cos\theta)$ curve.

It is interesting to note that for certain values of R and a the curve is very nearly a straight line, in agreement with the observations of Laby and Bingham(4). Under such conditions, however, the dispersion is comparatively small, so in the present experiments normality of spectra was sacrificed for high dispersion.

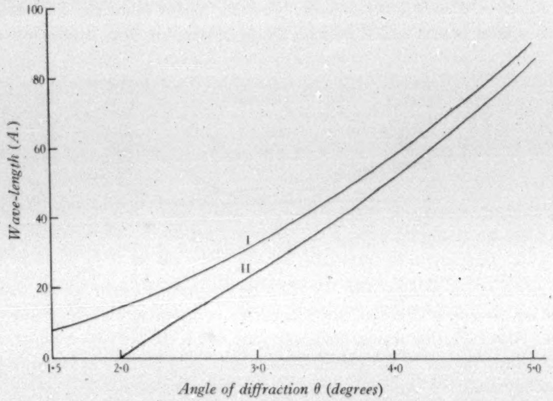


Figure 2. Dispersion of spectrograph. (I) Graph of $D(1-\cos\theta)$. (II) Graph of equation (3) when $i_0=\theta_0=2°$, $R=57{\cdot}3$, $a=1$.

θ_0 When θ is less than θ_0, where θ_0 is the angle of reflection, i.e. the value of θ for which $\lambda=0$, the curve becomes much steeper, and accordingly the dispersion much smaller. For this reason negative orders were not used in the present work.

Theory of the single-slit grating spectrometer: resolving-power. X-ray plane-grating spectrometers, unlike optical spectrometers in which lenses can be employed require a very narrow beam of rays, and this requirement limits the number of grating lines that can be used, and hence the resolving-power. The resolving-power of the plane grating, when treated as a Fraunhofer problem in diffraction, was first given by Rayleigh. Here, however, it must be treated as a Fresnel problem, for the source and image are at finite distances from the grating. Although an exact solution of this problem has not yet been given, Prins(3) has given an approximate theory, which will be slightly modified here to apply to the present instrument. The accuracy of Prins's method may be gauged by applying it to the case of the concave grating; it yields a value for the resolving-power which is about 20 per cent higher than that which Mack, Stehn and Edlen⁽²⁾ obtained by a stricter mathematical treatment. In addition to limiting the resolving-power, Porter⁽⁵⁾ has shown that the use of divergent beams with a plane grating introduces an asymmetrical structure into the spectral lines, so that their maxima are displaced slightly from the position they would occupy if parallel light were used.

In figure 3, if A is the source, A' a point on the photographic plate, O the centre of the grating, and P any other grating-element distant y from O, then the path difference between AOA' and APA' can be expressed as a power series in y, thus:

$$APA' - AOA' = y\,(\cos i - \cos\theta) + \tfrac{1}{2}y^2\,(\sin^2 i/r + \sin^2\theta/r') + \quad \ldots\ldots(4),$$

where i and θ are the glancing angles of incidence and diffraction respectively, and i, θ
r and r' are the distances of source and image respectively from O. r, r'

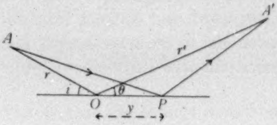


Figure 3.

The condition for the formation of a spectral line of wave-length λ is that the linear term should be an integral multiple of λ, when y is equal to one grating-space. This is the well-known relation

$$n\lambda = D\,(\cos i - \cos\theta).$$

Prins's criterion for best resolution is that the quadratic term should not exceed $\frac{1}{2}\lambda$ from the centre of the grating to the edge. So the optimum length of ruling, $2Y$, is that which satisfies the equation

$$\tfrac{1}{2}Y^2\,(\sin^2 i/r + \sin^2\theta/r') = \tfrac{1}{2}\lambda \qquad \ldots\ldots(5).$$

The maximum resolving-power is then the product of the number of rulings contained in this optimum length, and the spectral order

$$R = \frac{2Yn}{D} = \left[\frac{4n^2\lambda}{D^2\,(\sin^2 i/r + \sin^2\theta/r')}\right]^{\frac{1}{2}}$$

$$= \left[\frac{4n\,(\cos i - \cos\theta)}{D(\sin^2 i/r + \sin^2\theta/r')}\right]^{\frac{1}{2}}$$

$$= \left[\frac{2nr'\,(1 - i^2/\theta^2)}{D\,(1 + r'i^2/r\theta^2)}\right]^{\frac{1}{2}} \qquad \ldots\ldots(6),$$

approximately. In the present instrument, $r = 2$ cm., $r' = 50$ cm., $D = 23{,}520$ A. If $n = 1$ and $i^2/\theta^2 = 0{\cdot}1$ the resolving-power is 330. If $n = 1$ and $i^2/\theta^2 = 0{\cdot}01$ it is 640. The asymptotic value as i^2/θ^2 approaches 0 is 650. Now when $\lambda = 0$ we have normal reflection, and $i=\theta=i_0$. As λ increases i diminishes while θ increases, and the ratio i/θ diminishes from the value 1 the more rapidly the smaller the value of i_0. Hence the advantage of using small angles of incidence.

In actual practice the pairs of lines given in table 3 have been resolved. However, it is estimated that the lines could have been resolved in the first case if they had been about four times closer, and in the second and third cases if they had been about twice as close. So we may take the resolving-power actually realized to be about 200, which agrees at least in order of magnitude with the theoretical values.

Table 3

Pairs of lines	λ (A.)	$\delta\lambda$ (A.)	Resolution
Cu Lα I, Cu Lβ I	13·3	0·276	48
Pt Mα II, Zn Lα I	12·1	0·15	80
Pt Mγ I, S Kα I	5·3	0·052	102

The influence of the resolving-power on the width of a spectral line is most easily arrived at by considering the problem from the point of view of geometrical optics. Since, for a given slit-width and distance from slit to plate, the scale of the diffraction pattern is proportional to the wave-length, it is evident that in the case of X-rays, geometrical optics will give a sufficiently accurate representation of the formation of a line. This method has the advantage of giving a very clear conception of the variation of line-widths with wave-length for a given angle of incidence, and also the variation with angle of incidence for a given wave-length. These relations were first determined experimentally by measuring the width of the Cu Lα line in as many orders as possible on five typical spectra, one for each of the angles of incidence used. The results, figure 4, show that the reflected beam is in

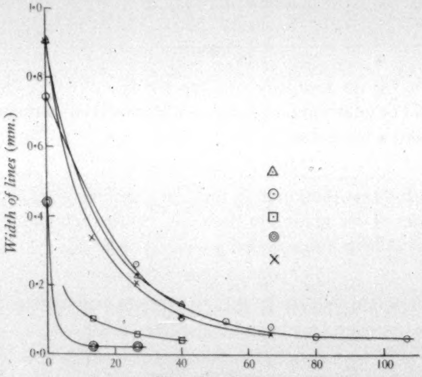


Figure 4. Variation of width of spectral lines with wave-length for various angles of incidence. △ $i_0=1·67^\circ$; ⊙ $i_0=1·55^\circ$; × $i_0=1·35^\circ$; ⊡ $i_0=0·4^\circ$; ◎ $i_0=0·5^\circ$.

every case very wide, but that the width of the diffracted lines decreases rapidly as the wave-length increases. The effect of the angle of incidence is not shown convincingly by the curves for films B 1, 9 and 14, as they lie close together and intersect one another. This is attributable to the fact that i_0 varies but little while the line-width depends upon other variable factors, such as the effective width of the focal spot and the degree of parallelism between the focal spot and the slit. The degree of resolution of the Zn Lα and Cu Lα lines in the first order, however, shows conclusively that the definition increases markedly as i_0 decreases, for when $i_0 = 1{\cdot}35^\circ$ these two lines are completely resolved; when $i_0 = 1{\cdot}55^\circ$ they are merged into a single wide line, and when $i_0 = 1{\cdot}67^\circ$, this line is but an ill-defined smudge.

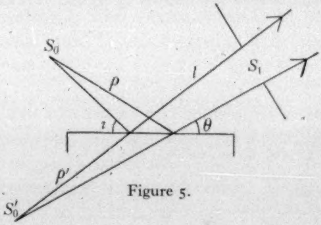


Figure 5.

The curves show definitely that the change in i_0 from $1{\cdot}35^\circ$ to $0{\cdot}4^\circ$ is accompanied by a marked diminution in line-widths. Film B 32 shows still further improvement, due in this case to a refinement in the method of adjusting the parallelism of the focal spot and the slit.

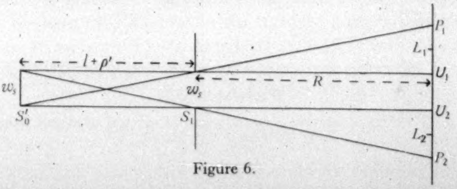


Figure 6.

The curves for two typical films, B 1 and B 24, will now be obtained theoretically. In figure 5 a divergent bundle falls on the grating from a point source S_0 at a distance ρ. Differentiation of equation (1) shows that the diffracted rays diverge ρ
from a virtual source S_0' at a distance ρ' behind the grating, where ρ'

$$\rho' = \rho \sin^2 \theta / \sin^2 i \qquad \text{......(7)}.$$

Actually the virtual source is not a point, but has a width of the same order of magnitude as the slit-width. For simplicity it will be assumed that source and slit have the same width w_s, figure 6. Then the spectral line will consist of an umbra w_s
U_1U_2 of width w_s and uniform intensity, and a penumbra P_1P_2 of total width $w_s[1 + 2R/(l+\rho')]$ in which the intensity falls off to zero at a uniform rate from U_1 to P_1 and from U_2 to P_2. We may take as a measure of the width of the line the

distance between points L_1 and L_2 at which the intensity has fallen to one-half of that in the umbra. Then, since L_1 and L_2 bisect P_1U_1 and P_2U_2, the line-width w will be given by

$$w = w_s\,[1 + R/(l+\rho')] \qquad \text{......(8).}$$

Now the values of i_0, r and g are known, where g is the value of MG, figure 1, when $i=\theta=i_0$. Then b is determined from $b/g=i_0$. For a series of values of θ the corresponding value of i is calculated from equation (2), and the corresponding wave-length from equation (1)—it is assumed that $b=c$, which was approximately true in practice; then ρ is given by $\rho=b/\sin i=b/i$ approximately, and ρ' is determined by equation (7). This enables w to be calculated for a series of wave-lengths. The results of these calculations are shown in figure 7. The curves show how, as the wave-length increases, the line-width in either case decreases rapidly at first and then approaches an asymptote whose value is the slit-width. It is also to be noted that the width of the reflected beam is always large, and is constant for a given slit-width, irrespective of i_0. For the smaller value of i_0, however, the initial drop in the curve is much more sudden, indicating much greater definition; in fact, for $i_0=0{\cdot}4^\circ$ the lines are fine (say $<0{\cdot}1$ mm.) from 2 A. upwards, while for $i_0=1{\cdot}55^\circ$ they are fine from about 25 A. upwards.

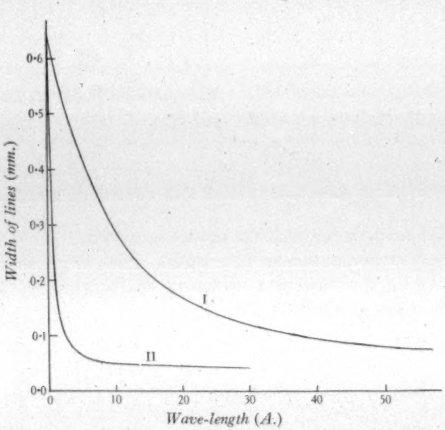


Figure 7. Theoretical relation between width of spectral lines and wave-length. (I) $i_0=1{\cdot}55^\circ$; (II) $i_0=0{\cdot}4^\circ$.

Though the curves have the same general form as those obtained experimentally, a quantitative agreement is not to be expected, for neither the projected width of the focal spot nor the geometrical constants of the apparatus (in particular the length b) are known with any high accuracy, and in addition the widths of the lines are affected markedly by the degree of parallelism between the focal spot and the slit. Thus the theoretical curve for $i_0=0{\cdot}5^\circ$ agrees much more satisfactorily with that obtained from film B 32 than with that obtained from film B 24. When the latter film was produced the adjustment of parallelism was made by an optical method, while in the case of the former film a more accurate photographic method was adopted.

It should be noted that the variation of line-width with angle of incidence and with wave-length is not a unique property of the present type of spectrograph but applies to all grating spectrographs, on account of the unavoidable divergence of the beams used. However, consideration of the function $\sin^2 \theta/\sin^2 i$ indicates that the variation with λ will not be so great in the case of an instrument employing two collimating slits, for in such a case the effect arises only through increase of θ with λ, while in the present instrument there is, in addition, an initially rapid decrease of i with increasing λ. The spectra obtained by the two-slit method (for instance by Witmer and Cork(6), and Siegbahn and Magnusson(7)) indicate a definite but small decrease in line-width as the distance of the line from the reflected line increases.

The variation of X-ray line-widths with angle of incidence appears to have been reported only by Witmer and Cork, who state that any diffracted line is broadened by the use of a large angle of incidence. Vonwiller(8), in Sydney, has recently observed the same phenomena in the optical region. Using a plane grating 2·7 cm. long, and with 14,431 lines to the inch, and a grazing angle of incidence of 1° 52′, he was able to resolve the sodium D doublet perfectly, even in the first order, when the collimator slit was 1·1 mm. wide; the direct and reflected beams were in this case about 2·5 mm. wide. Sharp lines were obtained with grazing angles of incidence even when the collimator lens was removed altogether, allowing a divergent beam to fall on the grating. Vonwiller states that the necessary condition for obtaining good spectral definition when the incident beam is divergent is that the ratio of θ to i should be large. This is in perfect agreement with the results obtained with all X-ray spectrographs.

§ 3. THE SPECTROGRAPH

The design of the single-slit spectrograph is indicated in figure 8. The vertical ring *A*, attached to a rigid base, carries the spectrometer on one side and the X-ray tube on the other. The outer cover *B* can be clamped to *A* without disturbing the spectrograph itself. The massive brass platform *C*, which supports the grating table, and the brass ring *D* which carries the slit, are both bolted firmly to *A*. To *D* is bolted a light-tight cone *E* of sheet-iron, which carries the plate-holder *F* at its farther end.

Each of the dismountable electrodes is attached to the spectrograph by means of a rubber gasket between metal discs which are clamped firmly together. To secure accurate registration of the target-face the gasket is sunk in a groove cut in one of the discs so that there may be metal-to-metal contact between the discs *G*.

To secure greater intensity in the spectra the electrodes are set obliquely to the axis of the spectrometer, bringing the focal spot to a distance of 18 mm. from the end of the ruling. The ruling extends for 19 mm. and the slit is placed 0·5 to 3·5 mm. beyond. The grazing angles of incidence have varied from 1·7° to 0·4°, while the slit-width was usually from about 25 μ. to 40 μ., but was sometimes reduced to 18 μ. or 12 μ.

The glass grating, with a grating-interval of 23,518 A., was ruled on the Grayson ruling engine in this University, and is the one previously used by Laby and Bingham(4). The slit is of the symmetrically closing type, the deviation of the centre amounting to only 4 μ. for slit-widths up to 125 μ.

Performance of X-ray tube. Oxide-coated platinum filaments were used in preference to tungsten, in order to minimize fogging of the plates by light and deposition of metal on the target. Although fogging by light was negligible, however, the deposit which accumulated on the face of the target during an exposure proved to be a serious drawback in the method, for on account of the small grazing angle of emission from the face of the target the soft X-rays would be largely absorbed in the deposit itself. Thus when a brass target was used, the copper and zinc L lines were often very faint, the strongest lines being those of platinum, barium, strontium and carbon.

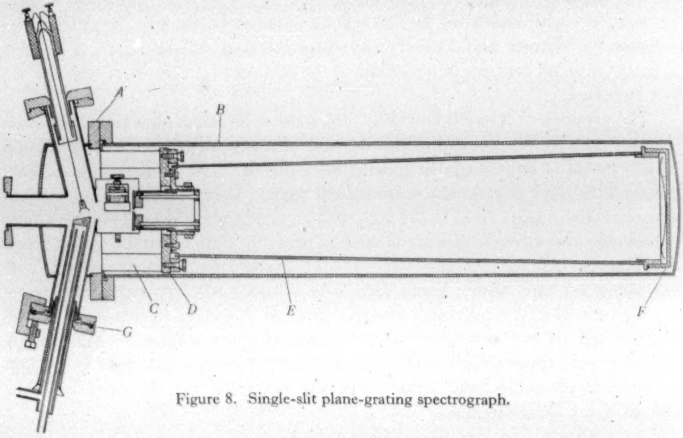


Figure 8. Single-slit plane-grating spectrograph.

If the thickness of the deposit increases at a uniform rate during the exposure, the intensity of the spectral lines from the target metal, instead of being proportional to the exposure, will approach an upper limit exponentially; so there is no advantage in continuing the exposure beyond a certain limit, which depends on how quickly the deposit is forming. It was found that the deposit was least when the tube was hardest, while the strongest target-metal lines were obtained when the exposures were very short, that is when conditions were such that these lines registered themselves before the deposit had time to form.

§ 4. WAVE-LENGTH MEASUREMENTS

The distances between spectral lines were measured by a projection method which has been in use for some years in this laboratory. The film or plate is placed on a glass stage which is propelled horizontally by the screw of a dividing-engine, and an image of the spectrum, magnified about eight times, is projected by an anastigmatic camera lens on to a fixed white screen on which is ruled a pair of fine parallel lines close together. Adjustment is made till the spectral line is exactly midway between these ruled lines, and the divided head of the engine is read to 0·001 mm.

In order to convert the measured distances on the film to wave-lengths an analytical expression was used to give a first approximation to the wave-length and the necessary corrections to this approximation were obtained from an empirical curve. The only assumption made in preparing this difference curve is that successive orders of one line are in the ratio 1:2:3.... This line, which may be called the calibrating-line, was selected for its homogeneity and its occurrence in numerous orders, and, since we are concerned only with relative determinations was given the arbitrary value of unity. Equation (3) may be written

$$n\lambda/D = 1 - \cos\theta - \delta. \qquad \delta$$

If λ_1 is the true wave-length of the calibrating-line and θ_1 its angle of diffraction in the first order, then λ_1, θ_1

$$\lambda_1/D = 1 - \cos\theta_1 - \delta_1. \qquad \delta_1$$

Table 4. Film B 24. Calibrating-line, copper $L\alpha_{12}$. $R = 254$ mm., $\theta = 0·4°$

Line	Order	Distance* from reflected line (mm.)	$\frac{1-\cos\theta}{1-\cos\theta_1}$	k	λ/λ_1 (i.e. x/n)
Cu $L\alpha_{12}$	1	6·987 ± 0·005	1·0000	0·0000	—
	2	10·553 ± 0·004	1·9797	0·0203	—
	3	13·293 ± 0·004	2·9571	0·0429	—
	4	15·620 ± 0·004	3·9410	0·0590	—
	5	17·663 ± 0·006	4·9206	0·0794	—
	7	21·219 ± 0·010	6·8843	0·1157	—
Zn $L\alpha_{12}$	1	6·629 ± 0·008	0·9200	−0·0019	0·9181
	2	10·044 ± 0·007	1·8196	0·0174	0·9185
	3	12·659 ± 0·006	2·7138	0·0360	0·9166
	4	14·882 ± 0·004	3·6138	0·0538	0·9169
	5	16·852 ± 0·008	4·5188	0·0722	0·9182
				Mean	0·9177 ± 0·0007
Cu $L\beta_1$	2	10·425 ± 0·006	1·9386	0·0199	0·9793
Zn $L\beta_1$	2	9·911 ± 0·003	1·7789	0·0165	0·8977
C $K\alpha$	1	14·107 ± 0·003	3·2853	0·0480	3·333
	2	20·849 ± 0·006	6·6650	0·1116	3·388
				Mean	3·360 ± 0·028
Pt $M\beta_1$	5	11·088 ± 0·005	2·1552	0·0223	0·4355
Pt $N_V N_{VI, VII}$	1	15·266 ± 0·008	3·7824	0·0578	3·8402

* The mean of ten readings, with the mean deviation from the mean.

Table 5. Film B 32. Calibrating-line, platinum M_{β_1}. $\theta = 0\cdot5°$

Line	Order	Distance from reflected line (mm.)	$\frac{1-\cos\theta}{1-\cos\theta_1}$	k	λ/λ_1 (i.e. x/n)
Pt $M\beta_1$	1	3·731 ± 0·001	1·0000	0·0000	—
	2	6·059 ± 0·002	1·9471	0·0529	—
	5	10·708 ± 0·085	4·7756	0·2244	—
Pt $M\alpha_{12}$	1	3·836 ± 0·001	1·0358	0·0020	1·0378
	2	6·207 ± 0·002	2·0179	0·0572	$1\cdot0375^5$
				Mean	1·0377 ± 0·0001
Sr $L\alpha_{12}$	1	4·211 ± 0·001	1·1696	0·0094	1·1790
Sr $L\beta_1$	1	4·103 ± 0·002	1·1301	0·0073	1·1374
Zn $L\alpha_{12}$	1	6·258 ± 0·003	2·0432	0·0586	2·1018
Cu $L\alpha_{12}$	1	6·613 ± 0·003	2·2188	0·0687	2·2875
	2	10·160 ± 0·003	4·3775	0·1991	2·2883
				Mean	2·2879 ± 0·0004
C $K\alpha$	1	13·734 ± 0·005	7·2874	0·3885	7·6759
	2	20·241 ± 0·008	14·4750	0·8784	7·6767
				Mean	7·6763 ± 0·0004
Pt $M_V N_{III}$	1	4·585 ± 0·003	1·3108	0·0172	1·3280

x Since δ is small, the ratio x of any line in any order to the calibrating-line in the first order may be written

k

$$x = n\lambda/\lambda_1 = (1-\cos\theta)/(1-\cos\theta_1) - k.$$

So $(1-\cos\theta)/(1-\cos\theta_1)$ was used as an analytical approximation, while for each order of the calibrating-line (for which $x = 1, 2, 3, \ldots$) the correction k was computed and plotted against $1-\cos\theta$.

The wave-lengths of the foregoing lines can be obtained either on the absolute grating scale or on the crystal scale, by taking their ratios to the $L\alpha_{12}$ line of copper and multiplying them by a measured value of the latter. As only two absolute grating measurements of $L\alpha_{12}$ of copper are available, and as they agree to only 1 in 400, the mean of the crystal determinations shown in table 6 is used in preference. These values are relative to the lattice constant for calcite.

Table 6

Observer	Wave-length of copper $L\alpha_{12}$ (A.)
Karlsson (9)	13·306
Siegbahn and Thoraeus (10)	13·316
Hjalmar (11)	13·31
	Mean 13·310

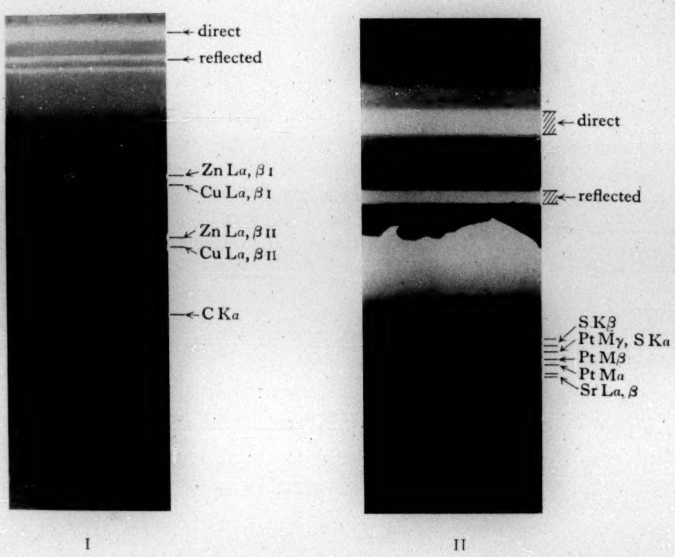

direct
reflected
Zn Lα, β I
Cu Lα, β I
Zn Lα, β II
Cu Lα, β II
C Kα
I
direct
reflected
S Kβ
Pt Mγ, S Kα
Pt Mβ
Pt Mα
Sr Lα, β
II

Figure 9. Typical spectra obtained with brass target. (I) $i=0{\cdot}4^\circ$, showing copper and zinc L lines. (II) $i=0{\cdot}5^\circ$, showing chiefly platinum, strontium and sulphur lines.

To face page 638

In table 7 are given the ratios found by the writer, the wave-lengths relative to $L\alpha_{12}$ of copper, which $=13{\cdot}310$ A., and previously existing crystal determinations.

Table 7

Line	Ratio	On crystal scale (A.)	Previous crystal determinations (A.)
Pt $M\alpha_1$	0·4535	6·037	6·034 Lindberg(12)
Pt $M\beta_1$	0·4362	5·808	5·816 ,,
Pt $M_V N_{III}$	0·5804	7·725	7·722 ,,
Pt $N_V N_{VI, VII}$	3·8402	51·116	None available
Sr $L\alpha_1$	0·5153	6·858	6·8486 Coster(13)
Sr $L\beta_1$	0·4971	6·617	6·6100 ,,
Cu $L\alpha_{12}$	(1·0000)	(13·310)	—
Cu $L\beta_1$	0·9793	13·035	13·027 Karlsson(9)
Zn $L\alpha_{12}$	0·9182	12·221	12·229 ,,
Zn $L\beta_1$	0·8977	11·949	11·960 ,,
C K	3·357	44·68	None available

Sources of error. As the measurements are entirely relative, the only sources of error are (1) in the measured distances on the films, and (2) in the calculation of the ratios. The chief error is that of setting the graticule of the measuring-instrument on the spectral lines measured. It has been found possible with practice, and by using as a graticule a pair of finely ruled parallel lines, to measure the distance between two well-defined spectral lines with a mean deviation of 2μ. or even 1μ. from the mean of ten readings. The mean deviation may amount to 5 or 10μ. if the lines are less well defined, or differ little in density from the background, even though they may be much narrower. There are two other sources of error: (i) The curve relating the photographic density of a line to the wave-length may not be symmetrical about its maximum. This asymmetry may be inherent in the radiation producing the line, or it may arise from some property of the spectrograph. Since the centre of gravity of the image may not then coincide with the position of the maximum of the density curve, the latter may be misjudged by visual observation. (ii) It has been found that in the drying of a film after development, the gelatine layer tends to alter its position slightly. To avoid this, Cooksey and Cooksey(14) recommend normalization of plates before use, but if care is taken to ensure evenness of drying, relative measurements should not be affected seriously. Celluloid films change in length when left *in vacuo* for some time owing presumably to the loss of some volatile constituent, but plates are free from this disadvantage.

As the intensity of the spectral lines usually decreased rapidly with increasing order, an accuracy of 1μ. or 2μ. was possible only in the first order of intense lines, the error increasing to about 10μ. in high orders. For this reason the accuracy of the above results is believed to range from 1 in 3000 for the best defined lines, to 1 in 1000 for the less well defined ones.

§ 5. STRUCTURE OF THE K LINE OF CARBON

The X-ray spectrograph, besides yielding wave-length determinations, can furnish information as to the width and structure of X-ray lines. Such information is of interest in the wave-length band under consideration, particularly in the K series of the elements in the first row of the periodic table, and the L series of the

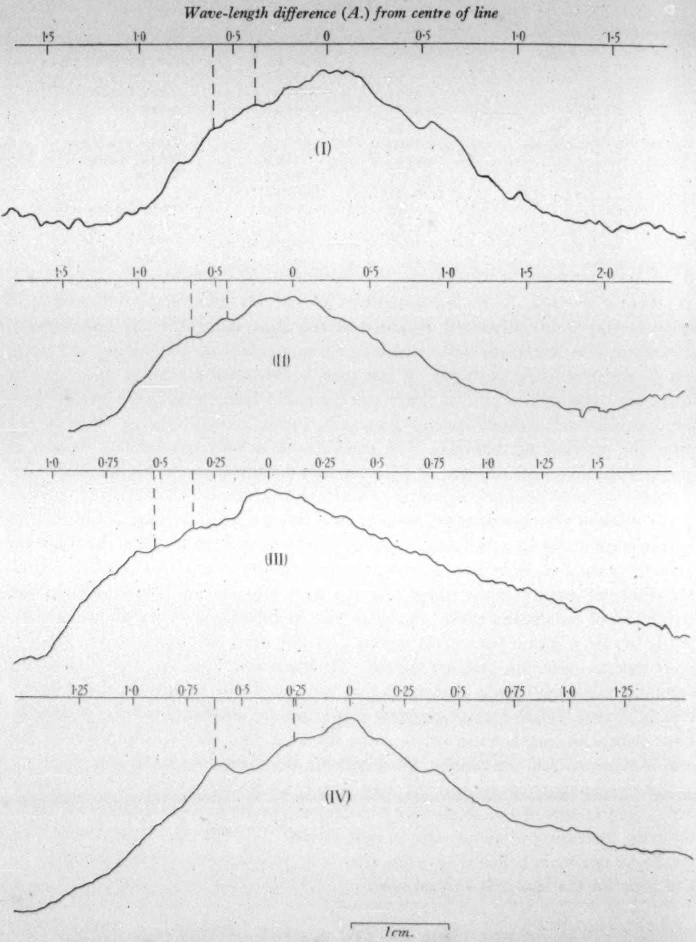


Figure 10. Microphotometer curves of the K line of carbon. (I) Order I, trace 1, $z=4{\cdot}17$; (II) order I, trace 2, $z=3{\cdot}77$; (III) order II, trace 1, $z=4{\cdot}00$; (IV) order II, trace 2, $z=3{\cdot}42$, where z is the distance (cm.) of galvanometer zero above the datum line.

elements in the second row, for diversity of opinion has existed as to the actual width of these lines (some of which are several angstroms wide), and as to whether or not they possess fine structure.

The most reliable work on the structure of the K series of the light elements has been done by Hautot(15) and Siegbahn and Magnussen(16), who obtained very high dispersion and resolving-power by using concave gratings of optimum width and of very large radius of curvature.

Hautot finds that the carbon K line consists of three components, each about 0·2 A. wide, which he interprets as being the normal α_1 line accompanied by satellites of shorter wave-length in accordance with Langer's theory. In the case of boron, Hautot shows that Langer's theory predicts no normal satellites, on account of the incompleteness of the 2p shell. He obtains, besides the α_1 line, an abnormal satellite of longer wave-length, 2·5 A. wide, whose intensity rises abruptly on the short-wave-length side and then falls off slowly towards the long-wave-length side. In the case of beryllium he finds a single line, 9 A. wide, with an abrupt short-wave-length edge. It is not the α_1 line, which cannot exist. To explain these two abnormal lines, each of great width and possessing a characteristic sharp edge on the short-wave-length side, Hautot suggests that a 2s electron is excited into some ill-defined metastable state, from which it can fall to the 1s level without infringing selection rules. The width of the normal diagram lines (about 0·2 to 0·3 A.) may be attributed to the fact that the upper energy level concerned in the transition is broadened by interaction with neighbouring atoms in the solid. In support of this statement, Hautot points out that he found the structure of the wide component of the boron spectrum to be independent of chemical combination, whereas Renninger(17) found the shape of the carbon line to vary with chemical combination. Further, Karlsson and Siegbahn(18) have shown that those lines of the K spectrum of aluminium and magnesium which arise from transitions from the M, or valence electron, shell are symmetrical when excited by the non-conducting oxide, but have sharp edges on the short-wave-length side when excited by the pure metal. Siegbahn and Magnusson(16) obtained the same effect in the L spectrum of aluminium.

Houston(19) has calculated the intensity-distribution in the carbon K line to be expected from (i) the free-electron model of Pauli(20) and Sommerfeld(21) and (ii) the bound-electron model of Bloch(22); but his results do not agree well with the experimental results of Hautot, Siegbahn and Magnusson, and Söderman(23). Yet in view of the work of Siegbahn and his co-workers, there is no doubt that the intensity-distribution in the normal lines is intimately connected with the conductivity of the solid emitting them, and it seems probable that an accumulation of more accurate intensity-distribution data may yield valuable information on the theory of the electrical conductivity of solids.

In the present experiments the carbon line often appeared in the spectra, arising as an impurity from hydrocarbon vapours which are invariably present in an apparatus in which grease and wax are used at joints. To obtain microphotometric curves of this line film B 24 was used, as it was the only film which combined good definition with a carbon-line density sufficient for the photometer. On this film the lines appeared in the first two orders, from each of which two traces were obtained on different portions of the line. The microphotometer consisted of a microscope with eye-piece removed, mounted horizontally. The film was placed on the stage of the microscope, and light from a point-o-lite arc was focused on it by the sub-stage condenser. The magnified image was received on an adjustable slit, behind which was the thermopile. Figure 10 shows a reproduction of the traces obtained. It is seen that on the short-wave side of the maximum there are two distinct components whose positions agree moderately well on the four curves. The irregularities on the long-wave side are not significant as they are not repeatable on the four curves.

From a knowledge of the dispersion at 45 A. and at 90 A., and of the magnification, 90, produced by the microphotometer, the distances of the two minor components from the maximum of the curve were determined with the results shown in table 8.

Table 8

Order	Trace	First component (A.)	Second component (A.)
I	1	0·38	0·60
I	2	0·41	0·66
II	1	0·35	0·53
II	2	0·26	0·63
	Means	0·35 ± 0·05	0·60 ± 0·04

This result agrees satisfactorily with that of Hautot, who obtained for the above separations on one occasion⁽²⁴⁾ 0·4 and 0·6 A. and on a later occasion⁽¹⁵⁾ 0·34 and 0·74 A., respectively. Siegbahn and Magnusson obtained only two maxima, 0·72 A. apart.

§ 6. ACKNOWLEDGMENTS

The author wishes to express his thanks to Prof. T. H. Laby, F.R.S., for designing the spectrometer in its first form, and for assistance in drafting the paper. Thanks are also due to Dr C. E. Eddy for his continued interest and helpful advice throughout the work.

REFERENCES

(1) Laby and Bingham. *Nature*, Lond., **126**, 915 (1930).
(2) Mack, Stehn and Edlen. *J. opt. Soc. Amer.* **22**, 245 (1932).
(3) Prins. *Z. Phys.* **69**, 618 (1931).
(4) Laby and Bingham. *Proc. roy. Soc.* **133**, 274 (1931).
(5) Porter. *Phil. Mag.* **5**, 1067 (1928).
(6) Witmer and Cork. *Phys. Rev.* **42**, 743 (1932).

(7) Siegbahn and Magnusson. *Z. Phys.* **62**, 435 (1930).
(8) Vonwiller. *J. roy. Soc. N.S.W.*, **66**, 486 (1932).
(9) Karlsson. *Ark. Mat. Astr. Fys.* **22**, 9 (1930).
(10) Siegbahn and Thoraeus. *J. opt. Soc. Amer.* **13**, 235 (1926).
(11) Hjalmar. *Phil. Mag.* **41**, 675 (1921).
(12) Lindberg. *Nova Acta Soc. Sci. upsal.* ser. iv, **7**, No. 7 (1931).
(13) Coster. *Phil. Mag.* **43**, 1070 (1922); **44**, 545 (1922).
(14) Cooksey and Cooksey. *Phys. Rev.* **36**, 80 (1930).
(15) Hautot. *J. Phys. Radium*, **5**, 20 (1934).
(16) Siegbahn and Magnusson. *Z. Phys.* **87**, 291 (1934).
(17) Renninger. *Z. Phys.* **78**, 510 (1932).
(18) Karlsson and Siegbahn. *Z. Phys.* **88**, 76 (1934).
(19) Houston. *Phys. Rev.* **38**, 1797 (1931).
(20) Pauli. *Z. Phys.* **41**, 81 (1927).
(21) Sommerfeld. *Z. Phys.* **47**, 1 (1928).
(22) Bloch. *Z. Phys.* **52**, 555 (1928).
(23) Söderman. *Z. Phys.* **65**, 656 (1930).
(24) Morand and Hautot. *C.R. Acad. Sci.*, Paris, **197**, 520 (1933).

621.395.623.5:534.213

SPHERICAL SOUND-WAVES OF FINITE AMPLITUDE

By N. W. McLACHLAN, D.Sc., M.I.E.E.

AND

A. L. MEYERS, B.Sc., A.M.I.E.E.

Communicated by Prof. A. O. Rankine, F.R.S., March 1, 1935. *Read in title May* 17, 1935

ABSTRACT. Owing to the use of acoustical outputs up to 500 watts in modern public-address loud-speaking apparatus, it is necessary to consider the distortion due (mainly) to the adiabatic pressure-volume relationship for air being non-linear. The physical principles underlying the transmission of sound-waves of finite amplitude in expanding waves are epitomized, and the differential equation for a horn of any cross-section is stated. The equation is solved for spherical-wave propagation, which includes the case of a conical horn having any solid angle up to 4π. Formulae are given which enable the shapes of the particle amplitude and pressure waves to be calculated to a second approximation. As the sound travels down the horn, the crests of the waves tend to overtake the troughs, and harmonics are created, the power associated therewith being transferred from the fundamental. The ratio of the power in the second harmonic to that in the fundamental is obtained, and comparison made with that for a uniform tube and an exponential horn. The analysis is illustrated by numerical examples, and design formulae are deduced from which the area of the horn-throat to keep the distortion below a prescribed limit can be calculated. Lamb's analysis for a uniform tube has been extended from the second to the third approximation.

§ 1. INTRODUCTION

The classical theory of sound formulated over half a century ago is based upon infinitesimal pressure-amplitudes. In modern loud-speaking apparatus, and in all musical instruments, such as the pedal organ, which are capable of generating high sound pressure, the amplitudes at the source are certainly not infinitesimal! The pressures employed in fog-signalling apparatus and in public-address loud-speakers are colossal from the point of view of simple theory. For example the pressure at the throat of a horn-type loud-speaker may be in the neighbourhood of 0·1 atmosphere, i.e. about 10^5 dyne/cm². Since a pressure of 5 dyne/cm². in the ear canal is perceived as quite a loud sound, the great strength of the modern acoustical source will be realized.

At any instant during the propagation of a plane sound-wave of finite amplitude, say in an infinite frictionless uniform tube, the density of the air varies from a maximum at a crest to a minimum at a trough, so the crests steadily gain on the troughs. This is shown in figure 1, which indicates a change in wave-form. The process is presumed to continue until the wave-slope becomes vertical. Beyond this point the mathematical analysis ceases to have any physical significance.

Since there is no expansion in a plane wave, the alteration in wave-form is much more marked in such a wave than it is in an expansive wave emitted from a horn or by a large conical diaphragm. The effect of expansion is a gradual reduction in pressure-amplitude, and this implies smaller variations in density and, therefore, less distortion than in a plane wave. The tendency in acoustical public-address work is to increase the power-output, and loud-speakers which radiate 500 watts(1) are now in use for directing operations under conditions where there is a good deal of background noise, e.g. in docks and at fires. It is possible that in the near future loud-speakers capable of radiating a kilowatt or more may be fitted to aeroplanes and used for "sky-shouting". In view of these developments, it seems desirable to investigate the creation of alien tones as the sound-waves expand from the source outwards.

In the present paper we start with a sinusoidal variation in particle-displacement with time, due to a moving diaphragm, and obtain formulae which enable the change in wave-form to be traced as the wave expands outwards. When the diaphragm-displacement comprises two or more simple harmonic motions of different frequencies, the analysis is more complicated than that given herein. Owing mainly to curvature of the $\{p, v\}$ relationship for air (i.e. $pv^\gamma = \text{const.}$) rectification occurs, so sum and difference frequencies are introduced. A similar effect occurs when a triode is operated on the non-linear portions of its characteristic. In both cases the analysis contains product terms whose factors represent sinusoidal oscillations of the same or of different frequencies. These products can be resolved into sum and difference components, e.g. $2 \cos \omega \cos \omega_0 = \cos(\omega + \omega_0) + \cos(\omega - \omega_0)$. In modern technical phraseology, one frequency component modulates another. Hence terms of the above nature are known as modulation products. When the amplitude at any part of a reproducing system, e.g. a microphone, amplifier, loud-speaker mechanism, or the air itself, exceeds a certain value, rectification ensues, extraneous frequencies are created and modulation products appear in the analysis.

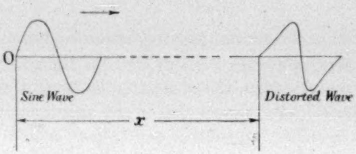


Figure 1. Diagram illustrating change in form of plane wave due to non-linearity of the medium.

No characteristic is truly linear, so that to avoid serious distortion of the above nature, when a very complex wave-form involving many different frequencies is reproduced, the amplitude must be limited. For high-quality reproduction from a loud-speaker, the limitation can be specified as follows: With a pure sine-wave input, the second harmonic due to curvature of the $\{p, v\}$ characteristic for air must be at least 30 decibels below the level of the fundamental at the normal listening position.

As a deduction from the foregoing, it should be possible to hear difference tones due to two or more intense supersonic sounds, provided the said tones lie within the audible frequency range.

§ 2. THE DIFFERENTIAL EQUATION AND ITS PHYSICAL INTERPRETATION

The equation for the propagation of sound-waves of finite amplitude in an extremely long, rigid, frictionless horn of any cross-section having a linear axis, is[(2)]

$$\xi'' + \frac{\chi'}{\chi}(1+\xi') = \frac{\ddot{\xi}}{c^2}(1+\xi')^{\gamma+1}\chi^{\gamma-1} \qquad \ldots\ldots(1),$$

ξ, x where ξ is the particle-displacement from abscissa x

χ $$\xi' = \partial\xi/\partial x,\ \xi'' = \partial^2\xi/\partial x^2,\ \ddot{\xi} = \partial^2\xi/\partial t^2,\ \chi' = d\chi/dx,$$

c, γ c is the velocity of sound for infinitesimal amplitudes, and $\gamma = 1{\cdot}4$ for air.

The simplest case is that of plane waves in a reflectionless uniform tube, when equation (1) degenerates to

$$\xi'' = \frac{\ddot{\xi}}{c^2}(1+\xi')^{\gamma+1} \qquad \ldots\ldots(2),$$

which has been solved to the second approximation by Lamb[(3)]. When expanded, the right-hand side of equation (2) contains terms of the second and higher orders in ξ, so the equation is non-linear. This is due to curvature of the $\{p, v\}$ relationship for air. Various writers[(3)] have shown that if the relationship were linear and of the form $p = a - bv$, equation (2) reduces to $\xi'' = \ddot{\xi}/c^2$, which being linear implies absence of distortion. This means that $\gamma = -1$ in equation (2).* If we put $\gamma = -1$ in equation (1), distortion due to the medium disappears, and the general horn equation becomes

$$\xi'' + \frac{\chi'}{\chi}(1+\xi') = \frac{\ddot{\xi}}{c^2\chi^2} \qquad \ldots\ldots(3).$$

ϕ Now from figure 2, $\chi = \phi(x+\xi)/\phi(x)$, where $\phi(x)$ is the cross-sectional area of the wave at abscissa x. Thus both χ and χ' contain terms in ξ, and equation (3) is non-linear, so its solution contains harmonics of the oscillation impressed at the

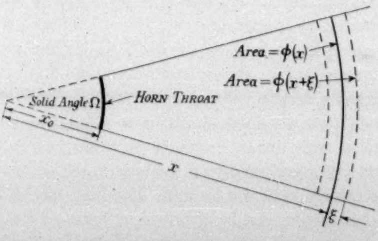


Figure 2. Diagram illustrating the propagation of a spherical wave: $\chi = \phi(x+\xi)/\phi(x)$. Area of horn-throat $A_0 = \Omega x_0^2$.

---

* It is not suggested that this holds for any known medium!

horn-throat. If we consider the air particles over any section at abscissa x to be replaced by a very thin diaphragm, its area alters progressively as it moves from end to end of its swing. Consequently the volumes swept out on either side of the mean position are unequal, and non-linear distortion ensues. In practice, therefore, there are two sources of distortion, namely (i) non-linearity of the medium, and (ii) variation in the area of the cross-section as described above. The expanding area of the wave-front causes the pressure to be attenuated with increase in abscissa x, and this partially neutralizes the distortion. Moreover, equation (1) consists, so to speak, of certain terms which promote distortion and others which alleviate it. This point is illustrated below.

For spherically expanding waves, including those propagated in a conical horn[2] of any solid angle Ω from 0 to 4π, $\chi=(1+\xi/x)^2$ and $\chi'=2(1+\xi/x)(\xi'/x-\xi/x^2)$. Ω Substituting these values in equation (1) and rejecting terms of higher orders in ξ than the second, we obtain

$$\frac{\chi'}{\chi}(1+\xi') \doteqdot \frac{2\xi'}{x}-\frac{2\xi}{x^2}+\frac{2}{x}\left(\xi'-\frac{\xi}{x}\right)^2 \qquad \text{......(4),}$$

and

$$(1+\xi')^{\gamma+1}\chi^{\gamma-1} \doteqdot 1+(\gamma+1)\xi'+2(\gamma-1)\frac{\xi}{x} \qquad \text{......(5).}$$

Substituting from equations (4) and (5) in equation (1), the differential equation for spherical wave-propagation, to the second order of terms in ξ, is

$$\xi''+\frac{2\xi'}{x}-\frac{2\xi}{x^2}-\frac{\ddot{\xi}}{c^2}=\frac{\ddot{\xi}}{c^2}\left[(\gamma+1)\xi'+2(\gamma-1)\frac{\xi}{x}\right]-\frac{2}{x}\left[\xi'-\frac{\xi}{x}\right]^2 \qquad \text{......(6).}$$

The various terms in equation (6) can be allocated as follows. Distortion due to the medium is represented by $\frac{\ddot{\xi}}{c^2}(\gamma+1)\xi'$; that due to the varying-area effect is represented by $-\frac{2}{x}\left[\xi'-\frac{\xi}{x}\right]^2$; whilst $\frac{2\ddot{\xi}}{c^2}(\gamma-1)\frac{\xi}{x}$ is a hybrid term involving both forms of distortion. These three are partially neutralized by $\left(\frac{2\xi'}{x}-\frac{2\xi}{x^2}\right)$, which is associated with attenuation due to expansion of the wave-front as it travels down the horn.

§ 3. SOLUTION OF EQUATION (6)

For the first approximation we take

$$\xi''+\frac{2\xi'}{x}-\frac{2\xi}{x^2}-\frac{\ddot{\xi}}{c^2}=0 \qquad \text{......(7),}$$

of which the solution for a diverging wave of frequency $\omega/2\pi$, in a horn devoid of reflection, is

$$\xi_1=A_1e^{i(\omega t-kx)}\left[\frac{ik}{x}+\frac{1}{x^2}\right] \qquad \text{......(8),}$$ ξ_1, ω, t

where A_1 is an arbitrary constant, $i=\sqrt{(-1)}$ and $k=\omega/c$, the distance phase constant. A_1, k, i

There is also a solution of equation (7) which is independent of t, and which x_0 vanishes at the horn-throat where $x=x_0$. The said solution* is

$$\xi=A\,(x-x_0)+B\left(\frac{1}{x^2}-\frac{1}{x_0{}^2}\right).$$

It represents a unidirectional displacement of the air particles which is akin to a rectification effect in a valve, whereby the anode feed current increases. This part of the solution is not required in the present analysis, where alternating components only are considered. Under the initial conditions to be imposed later, the constants A, B A and B are indeterminate and may have any value. This remark applies also to Lamb's solution for waves of finite amplitude in a frictionless uniform tube of unlimited extent.†

If the air particles at the horn-throat, where $x=x_0$, vibrate according to the law ξ_0 $\xi=\xi_0 e^{i\omega t}$, then from equation (8) $A_1=\xi_0 e^{ikx_0}/(ik/x_0+1/x_0{}^2)$, it being understood that the real part of equation (8) is to be taken ultimately. Using this value of A_1 in equation (8) we get

$$\xi_1=\frac{\xi_0 x_0{}^2}{x^2}\left[\frac{1+ikx}{1+ikx_0}\right]e^{i\theta} \qquad \ldots\ldots(9),$$

θ where

$$\theta=\{\omega t-k\,(x-x_0)\}.$$

The real part of equation (9) is

$$\xi_1=\frac{\xi_0 x_0{}^2}{(1+k^2x_0{}^2)}\left\{\left(\frac{k^2x_0}{x}+\frac{1}{x^2}\right)\cos\theta-\left(\frac{k}{x}-\frac{kx_0}{x^2}\right)\sin\theta\right\} \qquad \ldots\ldots(10).$$

From equation (10) we get, by differentiation with respect to x,

$$\xi_1{}'=\frac{\xi_0 x_0{}^2}{1+k^2x_0{}^2}\left\{\left(\frac{k^2}{x}-\frac{2k^2x_0}{x^2}-\frac{2}{x^3}\right)\cos\theta+\left(\frac{k^3x_0}{x}+\frac{2k}{x^2}-\frac{2kx_0}{x^3}\right)\sin\theta\right\} \ldots\ldots(11).$$

We also have

$$\frac{\ddot{\xi}}{c^2}=-k^2\xi_1 \qquad \ldots\ldots(12).$$

Substituting from equations (10), (11), (12) in the right-hand side of equation (6), we find, after some reduction, that provided the r.h.s. $\ll 1$ the second approximation to equation (1) is

$$\xi''+\frac{2\xi'}{x}-\frac{2\xi}{x^2}-\frac{\ddot{\xi}}{c^2}=\frac{-\xi_0{}^2x^4}{(1+k^2x_0{}^2)^2}\left\{\left[\frac{k^6(\gamma+1)\,x_0}{x^2}-\frac{\frac{1}{2}k^4(k^2x_0{}^2-1)\,(\gamma+7)}{x^3}-\frac{20k^4x_0}{x^4}\right.\right.$$
$$\left.+\frac{17k^2(k^2x_0{}^2-1)}{x^5}+\frac{36k^2x_0}{x^6}-\frac{9\,(k^2x_0{}^2-1)}{x^7}\right]\cos 2\theta+\left[\frac{\frac{1}{2}k^5(k^2x_0{}^2-1)\,(\gamma+1)}{x^2}\right.$$
$$\left.\left.+\frac{k^5x_0(\gamma+7)}{x^3}-\frac{10k^3(k^2x_0{}^2-1)}{x^4}-\frac{36k^3x_0}{x^5}+\frac{18k\,(k^2x_0{}^2-1)}{x^6}+\frac{18kx_0}{x^7}\right]\sin 2\theta\right\}\ldots(13).$$

Terms independent of t have been omitted from the right-hand side of equation (13), since they would introduce a unidirectional term, which is not needed, in the particular integral. Provided $kx>5$, terms of lower order than x^{-3} on the right-hand side of equation (13) can be neglected, so for the second approximation we finally obtain the equation

---

* This is also the solution to equation (1), when density is constant (reference (2), p. 199), as it would be for static displacement in a uniform tube.

† Reference (3), p. 183, equation (30).

$$\xi'' + \frac{2\xi'}{x} - \frac{2\xi}{x^2} - \frac{\ddot{\xi}}{c^2} = \xi_0{}^2 \left\{ \left(\frac{B_1}{x^2} + \frac{B_2}{x^3} \right) \cos 2\theta + \left(\frac{C_1}{x^2} + \frac{C_2}{x^3} \right) \sin 2\theta \right\} \quad \ldots\ldots(14),$$

where

$$B_1 = -\frac{k^6 x_0{}^5 (\gamma + 1)}{(1 + k^2 x_0{}^2)^2}; \quad C_1 = -\frac{\frac{1}{2} k^5 x_0{}^4 (k^2 x_0{}^2 - 1)(\gamma + 1)}{(1 + k^2 x_0{}^2)^2};$$

B_1, C_1

$$B_2 = -\frac{(\gamma + 7)}{k(\gamma + 1)} C_1; \quad C_2 = \frac{(\gamma + 7)}{k(\gamma + 1)} B_1.$$

B_2, C_2

The complementary function of equation (14) for the second harmonic, of frequency ω/π, transmitted down the horn is

$$\xi_{2c} = \left[\frac{2ik}{x} + \frac{1}{x^2} \right] e^{2i(\omega t - kx)} \quad \ldots\ldots(15),$$

ξ_{2c}

so we have now to find a particular integral of equation (14). It can be shown that the formula*

$$\xi_{2p} = \xi_{2c} \int^x \frac{dx}{x^2 \xi_{2c}{}^2} \left[\int^x x^2 f(x) \xi_{2c} \, dx \right] \quad \ldots\ldots(16),$$

ξ_{2p}

is applicable to equation (14), where $f(x)$ represents the right-hand side thereof. If ξ_2 is the complete solution of equation (14) for the second harmonic, then $\ddot{\xi}_2/c^2 = -4k^2 \xi_2$ and $\ddot{\xi}_{2p}/c^2 = -4k^2 \xi_{2p}$, so that in deriving equation (16) the terms involving these quantities cancel out. To perform the integrations indicated in equation (16), we write ξ_2

$$x^2 f(x) = \xi_0{}^2 \left\{ \left(D_1 + \frac{D_2}{x} \right) e^{2i\theta} + \left(E_1 + \frac{E_2}{x} \right) e^{-2i\theta} \right\} \quad \ldots\ldots(17),$$

where $D_1 = \frac{1}{2}(B_1 - iC_1)$; $E_1 = \frac{1}{2}(B_1 + iC_1)$; $D_2 = \frac{1}{2}(B_2 - iC_2)$; $E_2 = \frac{1}{2}(B_2 + iC_2)$; and we use the following integral formulae: D_1, E_1, D_2 E

$$\int^x \frac{e^{ax}}{x^n} dx = \frac{e^{ax}}{ax^n} \left[1 + \frac{n}{ax} + \frac{n(n+1)}{a^2 x^2} + \ldots \right] \quad \ldots\ldots(18),$$

$$\int^x \frac{e^{ax} \log x \, dx}{x^n} = \frac{e^{ax} \log x}{ax^n} \left[1 + \frac{n}{ax} + \frac{n(n+1)}{a^2 x^2} + \ldots \right]$$
$$- \frac{e^{ax}}{ax^n} \left[\frac{1}{ax} + \frac{n + (n+1)}{a^2 x^2} + \frac{n(n+1) + n(n+2) + (n+1)(n+2)}{a^3 x^3} + \ldots \right] \quad \ldots\ldots(19).$$

The D_1 and E_1 terms of equation (17) give, respectively, to order x^{-3},

$$\xi_0{}^2 D_1 \left[\frac{i \log x}{4kx} + \frac{1}{16k^2 x^2} (2 \log x + 3) - \frac{3i}{64k^3 x^3} \right] e^{2i\theta} \quad \ldots\ldots(20),$$

$$\xi_0{}^2 E_1 \left[\frac{-i \log x}{4kx} + \frac{1}{16k^2 x^2} (2 \log x + 3) + \frac{3i}{64k^3 x^3} \right] e^{-2i\theta} \quad \ldots\ldots(21).$$

The D_2 and E_2 terms give, exactly,

$$\xi_0{}^2 (-iD_2/4kx^2)\, e^{2i\theta} \quad \ldots\ldots(22),$$

$$\xi_0{}^2 (iE_2/4kx^2)\, e^{-2i\theta} \quad \ldots\ldots(23).$$

---

* Reference (4), p. 102.

The complete particular integral of equation (14), to the degree of approximation considered, is the sum of expressions (20) to (23) inclusive. Since $i\,(D_1e^{2i\theta}-E_1e^{-2i\theta})=C_1\cos 2\theta-B_1\sin 2\theta$, $(D_1e^{2i\theta}+E_1e^{-2i\theta})=B_1\cos 2\theta+C_1\sin 2\theta$, and likewise for terms having the subscript 2, we find that

$$\xi_{2p}=\xi_0^2\,[F_1(x)\cos 2\theta-F_2(x)\sin 2\theta] \qquad \text{......(24)},$$

$$=\xi_0^2R\,\{[F_1(x)+iF_2(x)]\,e^{2i\theta}\} \qquad \text{......(25)},$$

where, when 1·4 is substituted for γ,

$$F_1(x)=\frac{C_1\log x}{4kx}+\frac{B_1}{16k^2x^2}(2\log x-11)-\frac{3C_1}{64k^3x^3} \qquad \text{......(26)},$$

$$F_2(x)=\frac{B_1\log x}{4kx}-\frac{C_1}{16k^2x^2}(2\log x-11)-\frac{3B_1}{64k^3x^3} \qquad \text{......(27)},$$

R and R denotes that the real part is to be taken. The retention of terms from equation (13) beyond those in x^{-3} would modify the last terms of functions (26) and (27), and add terms of lower order. To equation (25) must be added a complementary function of equation (14), so that the sum is zero when $x=x_0$, since the condition at that point is $\xi=\xi_0\cos\omega t$, the fundamental frequency alone being represented.* The complementary function in question is a multiple of the real part of equation (15) and may be taken as

$$\xi_{2c}=R\left\{G_1\left[\frac{2ik}{x}+\frac{1}{x^2}\right]e^{2i\theta}\right\} \qquad \text{......(28)},$$

G_1 where G_1 is a constant to be determined. The second approximation which gives the amplitude of the second harmonic is, therefore, the sum of equations (25) and (28), so

$$\xi_2=R\left\{\xi_0^2\,[F_1(x)+iF_2(x_0)]\,e^{2i\theta}+G_1\left[\frac{2ik}{x}+\frac{1}{x^2}\right]e^{2i\theta}\right\} \qquad \text{......(29)}.$$

The condition $\xi_2=0$ at $x=x_0$ gives

$$G_1=-\xi_0^2\,[F_1(x_0)+iF_2(x_0)]\Big/\left[\frac{2ik}{x_0}+\frac{1}{x_0^2}\right] \qquad \text{......(30)}.$$

Hence the complete solution of equation (14), to the degree of approximation considered, is the sum of equations (10) and (29), namely

$$\begin{aligned}\xi&=\xi_1\,(\text{fundamental})+\xi_2\,(\text{second harmonic})\\&=\frac{\xi_0x_0^2}{x^2}R\left\{\left[\frac{1+ikx}{1+ikx_0}\right]e^{i\theta}\right\}+\xi_0^2R\left\{\left[F_1(x)+iF_2(x)\right.\right.\\&\qquad\left.\left.-\frac{x_0^2}{x^2}\left(\frac{1+2ikx}{1+2ikx_0}\right)(F_1(x_0)+iF_2(x_0))\right]e^{2i\theta}\right\}\end{aligned} \qquad \text{......(31)}.$$

§ 4. AMPLITUDE-RATIO OF SECOND HARMONIC TO FUNDAMENTAL

If we retain only the first terms in equations (26) and (27), when $kx>10$, and if in addition $k^2x_0^2\gg 1$, equation (31) reduces to

$$\xi\doteq\frac{\xi_0x_0}{x}\cos\theta-\xi_0^2\sqrt{\left(1+\frac{4}{k^2x_0^2}\right)}\cdot\frac{k^2x_0^2(\gamma+1)}{8x}\log\frac{x}{x_0}\cos(2\theta+\tan^{-1}2/kx_0) \quad \text{...(32)},$$

---

* As shown in § 5 this does not hold for the sound-pressure.

so that the ratio of the particle-amplitudes is

$$\frac{\xi_2}{\xi_1}=\xi_0\sqrt{(1+4/k^2x_0^2)}\,.\,\frac{k^2x_0\,(\gamma+1)}{8}\log\frac{x}{x_0} \qquad \ldots\ldots(33).$$

When x_0 is extremely large and $x/x_0 \rightarrow 1$, the portion of the cone between x_0 and x degenerates to a uniform tube. Putting $x=x_0+l$, we obtain l

$$\log(x/x_0)=\log(1+l/x_0),$$

and since $l/x_0 \ll 1$, $\log(x/x_0)\rightarrow l/x_0$. Hence if $\theta_1=(\omega-kl)$, equation (32) can be written θ_1

$$\xi \doteq \xi_0\cos\theta_1-\tfrac{1}{8}\,\xi_0^2k^2\,(\gamma+1)\,l\cos 2\theta \qquad \ldots\ldots(34),$$

which is identical with the formula for plane waves in a uniform tube, l being the distance from the source$^{(3)}$. It should be observed that equations (31), (32) and (33) are independent of Ω, the solid angle of the conical horn. This is only to be expected since the propagation is spherical and $A^{-1}.dA/dx$, the rate of expansion per unit area, is independent of Ω. In the case of a diaphragm vibrating at the throat of a conical horn, the analysis only applies if the amplitude of the air particles is normal to the portion of the sphere intercepted by the solid angle of the horn. This condition may be fulfilled if the diaphragm is larger than the throat and is coupled thereto by a throat chamber$^{(2)}$. If the diaphragm were flat and if its radius were equal to that of the throat, the sound would be propagated in the form of a beam at frequencies where the wave-length was of the same order or less than the size of the diaphragm$^{(5)}$.

Owing to the factor $\log(x/x_0)$ in equation (33), the amplitude-ratio of the second harmonic to the fundamental increases with increase in distance from the throat, ultimately becoming infinite with x, which seems to be out of the question. For all practical purposes it is probably safe to restrict equation (33) to the conditions given in §7. This puts a limit to the value of x. These conclusions also apply, *mutatis mutandis*, to the tube formula (34).

§ 5. FORMULA FOR THE SOUND-PRESSURE

From equation (17) from p. 202 of *Loud Speakers*$^{(2)}$, the sound pressure at *variable* abscissa $(x+\xi)$ is given, to the second order of terms in ξ, by

$$-p/\rho_0c^2 \doteq \left\{\xi'+\frac{2\xi}{x}-\frac{1}{2}(\gamma+1)\,\xi'^2-\frac{2\gamma\xi\xi'}{x}-(2\gamma+1)\frac{\xi^2}{x^2}\right\} \qquad \ldots\ldots(35),$$ p

ρ_0 being the density of the air in the absence of wave-motion. ρ_0

Differentiating equation (31) with respect to x we get

$$\begin{aligned}\xi'=\frac{\xi_0x_0^2}{x^3}R\left\{\left[\frac{k^2x^2-2-2ikx}{1+ikx_0}\right]e^{i\theta}\right\}&+\xi_0^2R\left\{[F_1'(x)+iF_2'(x)]\,e^{2i\theta}\right\}\\ &-\xi_0^2R\left\{2ik\,[F_1(x)+iF_2(x)]\,e^{2i\theta}\right\}\\ &-\frac{2\xi_0^2x_0^2}{x^3}R\left\{\left[\frac{2k^2x^2-1-2ikx}{1+2ikx_0}\right][F_1(x_0)+iF_2(x_0)]\,e^{2i\theta}\right\} \qquad \ldots\ldots(36).\end{aligned}$$

Inserting equations (31) and (36) in the first two terms of equation (35), we get

$$\frac{\xi_0 k^2 x_0^2}{x} R\left\{\frac{e^{i\theta}}{1+ikx_0}\right\}+\xi_0^2 R\{[F_1'(x)+iF_2'(x)]\,e^{2i\theta}\}$$
$$+\frac{2\xi_0^2}{x} R\{[1-ikx]\,[F_1(x)+iF_2(x)]\,e^{2i\theta}\}$$
$$-\frac{4\xi_0^2 k^2 x_0^2}{x} R\left\{\left[\frac{F_1(x_0)+iF_2(x_0)}{1+2ikx_0}\right]e^{2i\theta}\right\} \quad \ldots\ldots(37).$$

In addition to expression (37), the last three terms of equation (35) produce from the fundamental a second-harmonic pressure and a unidirectional component. Neglecting the latter, we have for these terms

$$-\tfrac{1}{2}(\gamma+1)\frac{\xi_0^2 x^4}{2x^6} R\left\{\left[\frac{k^2x^2-2-2ikx}{1+ikx_0}\right]^2 e^{2i\theta}\right\}$$
$$-\frac{2\gamma}{x}\cdot\frac{\xi_0^2 x_0^4}{2x^5} R\left\{\left[\frac{(1+ikx)\,(k^2x^2-2-2ikx)}{(1+ikx_0)^2}\right]e^{2i\theta}\right\}$$
$$-\frac{(2\gamma+1)}{x^2}\cdot\frac{\xi_0^2 x_0^4}{2x^4} R\left\{\left[\frac{1+ikx}{1+ikx_0}\right]^2 e^{2i\theta}\right\} \quad \ldots\ldots(38)$$
$$=-\frac{\xi_0^2 x_0^4}{2x^6} R\left\{\left[\frac{F_3(x)}{(1+ikx_0)^2}\right]e^{2i\theta}\right\} \quad \ldots\ldots(39),$$

F_3 where $F_3(x)=\tfrac{1}{2}(\gamma+1)\,k^4x^4-2ik^3x^3-5k^2x^2+6ikx+3$.

The total sound-pressure due to the fundamental and the second harmonic is $-\rho_0 c^2$ times the sum of expressions (37) and (39). The latter is of order $\xi_0^2 k^2 x_0^2/x^2$, and must be included if more than the leading terms of the second harmonic in equation (37) are retained. From equations (26) and (27) we obtain by differentiation

$$F_1'(x)=\frac{C_1}{4kx^2}(1-\log x)+\frac{B_1}{4k^2x^3}(6-\log x)+\frac{9C_1}{64k^3x^4} \quad \ldots\ldots(40),$$

and

$$F_2'(x)=\frac{B_1}{4kx^2}(1-\log x)-\frac{C_1}{4k^2x^3}(6-\log x)+\frac{9B_1}{64k^3x^4} \quad \ldots\ldots(41),$$

which for large values of kx are of lower order than $kF_1(x)$ and $kF_2(x)$. Hence when $kx \gg 1$, in compliance with the condition that $kx > 5$, stipulated in § 2, we can neglect $F_1'(x)$, $F_2'(x)$, and retain only the first terms of $F_1(x)$ and $F_2(x)$. With these provisos, expression (37) reduces to

$$p/\rho_0 c^2 \doteqdot -\frac{\xi_0 k^2 x_0^2}{x} R\left\{\frac{e^{i\theta}}{1+ikx_0}\right\}-\xi_0^2\frac{\log x}{2x} R\{(B_1-iC_1)\,e^{2i\theta}\}$$
$$+\frac{4\xi_0^2 k^2 x_0^2}{x} R\left\{\left[\frac{F_1(x_0)+iF_2(x_0)}{1+2ikx_0}\right]e^{2i\theta}\right\} \quad \ldots\ldots(42)$$
$$=-\frac{\xi_0 k^2 x_0^2}{x\sqrt{(1+k^2x_0^2)}}\sin(\theta-\epsilon_1)+\frac{\xi_0^2 k^5 x_0^4(\gamma+1)\log x}{4x(1+k^2x_0^2)}\sin(2\theta+\epsilon_2)$$
$$-\frac{4\xi_0^2 k^2 x_0^2}{x}\sqrt{\left(\frac{F_1^2(x_0)+F_2^2(x_0)}{1+4k^2x_0^2}\right)}\sin(2\theta+\epsilon_3) \quad \ldots\ldots(43),$$

where $\epsilon_1 = \tan^{-1} 1/kx_0$, $\epsilon_2 = \tan^{-1} 2kx_0/(k^2x_0^2 - 1)$, and $\epsilon_3 = \tan^{-1}[-F_2(x_0)/-F_1(x_0)] +$ $\epsilon_1, \epsilon_2, \epsilon_3$
$\tan^{-1} 1/2kx_0$, the negative signs indicating that the first quadrant is to be chosen.

The formulae in this section refer to the pressures at variable abscissa $(x+\xi)$, but when x is large enough ξ is sufficiently small for them to give the values at abscissa x. The value* p_x at x is $p - \xi \partial p/\partial x$, so the preceding condition implies p_x
(correctly) that $\xi \partial p/\partial x$ is negligible.

By putting $x = x_0$ in equations (37) and (39) it can be shown that there is a second-harmonic pressure at the throat, although it is very small compared with that due to the fundamental. This holds for a uniform tube and for any type of horn. Moreover in measuring the fundamental pressure at the throat, with a small side tube and a microphone, it is advisable to filter out harmonics from the meter input. There are no harmonics of the particle-velocity at $x = x_0$, since by hypothesis it there follows a sine law. We may remark that in any expanding horn the second-harmonic pressure per unit area at the throat will exceed that at abscissa x, provided x is large enough. This follows from the expansion of wave-front area, although, of course, the *total* second-harmonic pressure over that area increases as the wave travels outwards. Moreover, in attempting any rough analytical approximations, it would be incorrect to take $p_2 = 0$ at $x = x_0$ as a boundary condition.

The second-order or product terms in equation (35), which are represented by equation (39), indicate that part of the second harmonic is created by modulation of the main portion of the harmonic by the fundamental.

§6. FORMULA FOR THE PARTICLE-VELOCITY

Differentiating equation (31) with respect to t we have

$$\dot{\xi} = \frac{\xi_0 x^2}{x} R\left\{ i\omega \left[\frac{1+ikx}{1+ikx_0}\right] e^{i\theta}\right\} + \xi_0^2 R\{2i\omega\,[F_1(x) + iF_2(x)]\, e^{2i\theta}\}$$
$$- \frac{\xi_0^2 x_0^2}{x^2} R\left\{2i\omega \left[\frac{1+2ikx}{1+2ikx_0}\right] [F_1(x_0) + iF_2(x_0)]\, e^{2i\theta}\right\} \quad \text{......(44)}.$$

When $kx \gg 1$, equation (44) can be written

$$\dot{\xi} = p/\rho_0 c \quad \text{......(45)},$$

the value of p being given by equation (42) or (43). Thus the sound-pressure and particle-velocity in both the fundamental and the second harmonic ultimately fall into phase. Formula (44) applies to abscissa $(x+\xi)$, but when x is large enough it holds for abscissa x, as is explained in § 5.

§7. POWER-RATIO OF SECOND HARMONIC TO FUNDAMENTAL

In evaluating the power transmitted down the horn, we imagine the air particles over any section to be represented by a massless radially vibrating spherical cap whose mean position is abscissa x but whose variable area A is $\Omega(x+\xi)^2$, ξ being A
given by equation (31). The pressure on the cap is obtained from § 5, and the

---

* Reference (2), p. 203.

particle-velocity from § 6. When x is large enough $A \doteqdot \Omega x^2$, and the power is given by the mean value of $p\dot{\xi}\Omega x^2$ over a complete cycle of the fundamental, p and $\dot{\xi}$ being obtained from equations (43) and (45), respectively. Thus the total power

$$P_1+P_2=\tfrac{1}{2}\,(p_1\dot{\xi}_1+p_2\dot{\xi}_2)\,\Omega x^2=\left(\frac{p_1{}^2+p_2{}^2}{2\rho_0 c}\right)\Omega x^2 \qquad \text{......(46)},$$

p_1, p_2

by virtue of equation (45). The values of p_1 and p_2 in equation (46) are maximum pressures, so the factor $\frac{1}{2}$ must be used to obtain the mean square value. To simplify the succeeding analysis we can consider the case where $kx_0 \geqslant 1$. Formula (43) then gives for the maximum pressures

$$p_1=\xi_0\rho_0 c\omega x_0/x \qquad \text{......(47)},$$

and

$$p_2=\frac{\xi_0{}^2\rho_0 c\omega k^2 x_0{}^2\,(\gamma+1)}{4x}\log\frac{x}{x_0} \qquad \text{......(48)}.$$

P_1
P_2

From equation (46) the powers associated with these two components are P_1 and P_2, where

$$P_1=\tfrac{1}{2}\xi_0{}^2\rho_0 cA_0\omega^2\times 10^{-7} \qquad \text{......(49)},$$

and

$$P_2=\tfrac{1}{32}\xi_0{}^4\rho_0 cA_0\omega^2 k^4 x_0{}^2\,(\gamma+1)^2\,[\log\,(x/x_0)]^2\times 10^{-7} \qquad \text{......(50)},$$

A_0

where $A_0=\Omega x_0{}^2$, the throat-area, and the power is expressed in watts. Thus from equations (49) and (50)

$$\frac{P_2}{P_1}=\frac{\xi_0{}^2k^4x_0{}^2\,(\gamma+1)^2}{16}\left[\log\frac{x}{x_0}\right]^2=\frac{10^7 P_1\omega^2\,(\gamma+1)^2 x_0{}^2}{8\rho_0 c^5 A_0}\left[\log\frac{x}{x_0}\right]^2 \text{......(51)},$$

P_1 being the power in watts supplied to the horn-throat at frequency $\omega/2\pi$. Since this is given by equation (49), it appears from equation (50) that the total power increases with increase in distance from the throat. This apparent paradox is due to neglect in the foregoing analysis of the transfer of power from the fundamental to the harmonic, in accordance with the condition that the total power passing any section of the horn is the same as that supplied to the throat by the driving mechanism. From equation (51) it appears that the ratio P_2/P_1 increases indefinitely with x. The analysis is only tenable, however, if $P_2/P_1 \ngtr 0{\cdot}1$ and $\xi_2/\xi_1 \ngtr 0{\cdot}25$. It will be seen that the ratio in equation (51) could have been obtained from § 4 by taking $\dot{\xi}_2{}^2/\dot{\xi}_1{}^2=P_2/P_1=4\xi_2{}^2/\xi_1{}^2$, maximum values being implied. But this procedure is only valid if the phase angles between the pressure and particle-velocity are identical for each component frequency. It was necessary, therefore, to prove this first, as has been done in §§ 5 and 6. At the same time the result might have been inferred from the fact that in any horn the pressure and particle-velocity fall into phase as the distance from the source increases, since the inertia component is then negligible.

§ 8. INFLUENCE OF EXPANSION IN REDUCING SECOND HARMONIC

To illustrate reduction in the second harmonic due to expansion of the wave as it travels outwards from the horn-throat at $x=x_0$, we make comparison with propagation in an infinite frictionless tube. The ratio P_2/P_1 for the tube is obtained from equation (51) by the method used to find equation (34). Thus we obtain

$$\frac{P_2}{P_1}\,(\text{tube}) = \xi_0^2 k^4 (\gamma+1)^2\, l^2/16 \qquad \text{......(52)}$$

$$= 10^7 P_1 \omega^2 (\gamma+1)^2\, l^2/8\rho_0 c^5 A_0 \qquad \text{......(53),}$$

where $l=(x-x_0)$ and is the length of the tube over which distortion occurs, and P_1 is the input power in watts. Taking the ratio of equation (51) to equation (53), we get l

$$\varphi = \frac{\text{Second-harmonic power in conical horn}}{\text{Second-harmonic power in tube}} = \frac{x_0^2}{l^2}\left[\log \frac{x}{x_0}\right]^2 \qquad \text{......(54).}$$ φ

If $x_0/x \ll 1$, $l \doteqdot x$ and φ depends entirely upon x_0/x; so the smaller x_0 for a given horn length, the lower is the level of the harmonic in the horn compared with that in the tube. In using the above formulae the conditions $kx \gg 1$ and $k^2x_0^2 \gg 1$ must be observed. It is interesting to compare equation (51) with the value obtained some time ago by Goldstein and McLachlan for an exponential horn(6, 7, 8). For this case

$$P_2/P_1 \doteqdot \xi_0^2 k^4 (\gamma+1)^2/4\beta^2 = 10^7 P_1 \omega^2 (\gamma+1)^2/2\rho_0 c^5 \beta^2 A_0 \qquad \text{......(55),}$$

provided $k \gg \tfrac{1}{2}\beta$ and x is so large that $e^{-\frac{1}{2}\beta x} \ll 1$, β being the flaring-index in the formula $A = A_0 e^{\beta x}$, and P_1 the input power in watts. The ratio of equation (55) to equation (52) is β

$$\frac{\text{Second-harmonic power in exponential horn}}{\text{Second-harmonic power in conical horn}} = \frac{4}{\beta^2 x_0^2\,[\log (x/x_0)]^2} \quad \text{...(56).}$$

Which of the two horns gives the lesser distortion depends upon circumstances, i.e. upon the values assigned to the quantities in equation (56). For example a small flaring-index in an exponential horn entails a relatively long tube-like portion from the throat, and this would cause the second harmonic to exceed that in a portion of a conical horn of equal length. Beyond this point the exponential horn expands so rapidly that the second harmonic substantially reaches its final value at the end of the horn (if the horn be not less than 400 cm. long). On the other hand the harmonic in the conical horn continues to grow relatively to the fundamental, owing to the rate of expansion being then far below that of the exponential horn.

§ 9. NUMERICAL EXAMPLES

To illustrate the analysis, we take the case of a simple source set flush with the ground and radiating 3 kilowatts at 500 c./sec. Assume the source to be equivalent to a pulsating hemisphere whose radius $x_0 = 30$ cm. Then on the assumption of zero loss during transmission, the power-level of the second harmonic relative to that of the fundamental can be calculated. Substituting for the various quantities in equation (51), taking $x = 6 \times 10^4$ cm. and the area $A_0 = 2\pi x_0^2$, we obtain $P_2/P_1 = 0{\cdot}034$. Thus at 600 metres from the source the level of the second harmonic is $10 \log_{10} P_2/P_1$ or 15 decibels *below* that of the fundamental. The corresponding pressure-ratio p_2/p_1 is 0·18.

As another example we may give the condition for the level of the second harmonic in a conical horn to be not less than 30 decibels below that of the fundamental. Since this corresponds to a power-ratio of 10^{-3}, we find from equation (51) that the throat-area

$$A_0 > 10^{10} P_1 \omega^2 (\gamma+1)^2 x_0^2 [\log (x/x_0)]^2 / 8\rho_0 c^5 \qquad \ldots\ldots(57).$$

For an exponential horn under identical conditions we have, using the provisos below equation (55),

$$A_0 \geqslant 10^{10} P_1 \omega^2 (\gamma+1)^2 / 2\rho_0 c^5 \beta^2 \qquad \ldots\ldots(58),$$

or

$$A_0 \geqslant 0{\cdot}148 P_1 (n/n_c)^2 \qquad \ldots\ldots(59),$$

n, n_c where n is the transmission frequency and n_c the cut-off frequency of the horn, i.e. $\beta c/4\pi$, and P_1 is in watts.

APPENDIX

As a matter of interest we have extended Lamb's solution⁽³⁾ of equation (2) for plane waves in a uniform tube to the third approximation, and find that

$$\xi \doteqdot \xi_0 \left[1 - \frac{\xi_0^2 k^4 (\gamma+1)^2 l^2}{32}\right] \cos\theta_1 - \frac{\xi_0^3 k^3 (\gamma+1) l}{16} \sin\theta_1 - \frac{\xi_0^2 k^2 (\gamma+1) l}{8} \cos 2\theta$$
$$+ \frac{\xi_0^3 k^4 (\gamma+1)^2 l^2}{32} \cos 3\theta - \frac{\xi_0^3 k^3 \gamma (\gamma+1) l}{48} \sin 3\theta \qquad \ldots\ldots(60).$$

This should be compared with the second approximation given by equation (34).

REFERENCES

(1) *Wireless World*, **35**, 525 (1934).
(2) McLachlan. *Loud Speakers*, p. 200, equation (6) (1934).
(3) Lamb. *Dynamical Theory of Sound* (1925).
(4) Forsyth. *A Treatise on Differential Equations* (1933).
(5) McLachlan. *Bessel Functions for Engineers*, p. xi (1934).
(6) Goldstein and McLachlan. *Wireless Engr*, **11**, 27 (1934).
(7) —— —— *Wireless Engr*, **11**, 423 (1934).
(8) —— —— *J. Acoust. Soc. Amer.* **6**, 275 (1935).

541.135.22:532.13

THE TEMPERATURE VARIATION OF THE VISCOSITY OF AQUEOUS SOLUTIONS OF STRONG ELECTROLYTES

By W. J. SULSTON, B.Sc., A.Inst.P., Queen Mary College,
University of London

Communicated by Prof. H. R. Robinson, F.R.S., March 1, 1935. *Read in title May* 17, 1935.

ABSTRACT. The viscosity of aqueous solutions of potassium chloride and of potassium sulphate has been measured over a wide range of temperature. Every care has been taken to eliminate, or to correct for, sources of error. The results have been analysed by plotting $\frac{\phi/\phi_0-1}{\sqrt{c}}$ against $\sqrt{c}$, and in those cases where the relation between these variables proved to be linear the values of A and B in the equation

$$\frac{\phi}{\phi_0}=1+A\sqrt{c}+Bc$$

at each temperature have been calculated.

§ 1. INTRODUCTION

Until 1929 accurate information was lacking concerning the viscosity of electrolytes at low concentrations and at temperatures above 35° C. Since that date a considerable amount of work has been done at very low concentrations. An excellent summary both of early work and of the accurate measurements made between 1929 and 1933 appears in a paper by Jones and Talley(1). The formula, at first tentatively proposed by Jones and Dole(2),

$$\phi/\phi_0=1+A\sqrt{c}+Bc,$$

where ϕ is the fluidity of solution at given temperature θ° C., ϕ_0 is the fluidity of solvent at the same temperature, and c is the concentration of the solution, has since been confirmed for certain salts over limited ranges of concentration. For potassium chloride Joy and Wolfenden(3) showed that the equation was valid up to a concentration of about $0{\cdot}15N$ at 18° C., and up to a concentration of about $0{\cdot}5N$ at 35° C.

ϕ, θ, ϕ_0
c

Jones and Dole(2) had previously found the equation to hold for barium chloride at 25° C. Hood and Hohlfelder(4), however, found that the equation did not apply to the same salt at 18° C.

The Debye-Hückel theory has been applied to the problem of the viscosity of strong electrolytes. The general problem of any electrolyte solution has been solved by H. Falkenhagen and E. L. Vernon(5). The predictions of the theory have been confirmed for a number of salts by various workers, including Joy and Wolfenden, and Jones and Talley. Recently, however, Cox and Wolfenden(6), who used a new and extremely accurate method of measurement, obtained results for the values of A for magnesium sulphate and lanthanum sulphate which differ widely from those predicted by the theory.

The purpose of the present work is to measure the relative viscosity of potassium chloride over a range of temperature from 18° to 85° C. The relative viscosity of an aqueous solution of a salt is defined as the ratio of the viscosity of the solution to the viscosity of water at the same temperature. In view of the lack of data for unidivalent salts, measurements have also been made on potassium sulphate.

§ 2. METHOD OF EXPERIMENT

Viscometer. On account of the small difference between the viscosity of a dilute aqueous solution and that of water an accurate method of measurement was essential. The general design of the viscometer followed that given by Washburn and Williams(7). A somewhat wider capillary was, however, employed, and a kinetic-energy correction was applied. The instrument was made of pyrex glass and its approximate dimensions were:

Capacity of upper bulb	16 cm.^{3}
Length of capillary	20 cm.
Diameter of capillary	0·06 cm.
Diameter of lower bulb	8·0 cm.
Vertical diameter of lower bulb ...	4·5 cm.

The cleaning-solutions were found to have little or no action upon the glass.

In view of the range of temperature covered, it was considered undesirable to use an all-glass fitting for the manipulation of the liquid in the viscometer. Ground-glass joints were, however, employed to make connexion to the limbs of the instrument. The two limbs were normally connected by a three-way stopcock so that loss of liquid by evaporation was minimized. The liquid was drawn into the upper bulb before the commencement of a run by connecting that limb to a large and partially evacuated air reservoir.

In order to obtain comparable times of flow for different liquids it is essential that the viscometer should be placed at the same inclination to the vertical in all cases. The instrument was permanently attached to a rigid brass stand which could be screwed on to a plate fixed in the thermostat. The use of three pegs to bring the instrument to a standard position, as used by Appleby(8) and other workers, was adopted.

Thermostat and temperature-regulation. It is well known that the viscosity of water varies rapidly with temperature. In order to measure the viscosity of water at normal temperatures to an accuracy of 0·1 per cent it is necessary to maintain and to measure the temperature to 0·04° C.; hence it is essential to employ an efficient thermostat. A copper vessel of capacity about 40 litres and depth 40 cm. was used. It was well lagged with glass wool on all sides with the exception of a small glass window at the front through which the upper bulb of the viscometer was observed.

The bath was filled with water which was covered with a layer of oil to minimize evaporation at high temperatures.

Now while it is comparatively easy to design a thermostat to keep constant to ±0·002° C. at 25° or 35° C., it was necessary in the present work to cover a range from 18° to 85° C., and to be able to change from one temperature to a higher one fairly rapidly. It was finally decided to use a mercury-toluene regulator which operated a thermionic-valve-actuated relay. The regulator was of conventional design in the form of a helix, but to obviate the necessity of removing mercury, as the temperature was raised, by means of a pipette or some special device, it was fitted with a side tube about 3 cm. below the point at which contact was made. This tube was connected, through a stopcock, to a reservoir which could be moved vertically. As the temperature was raised in stages of 5° C., mercury could be removed from the regulator by opening the stopcock and lowering the reservoir. This device was used to obtain a rough adjustment of the amount of mercury in the regulator; a fine adjustment to set the temperature of the bath accurately was provided by soldering the contact with the mercury to a screw which moved vertically in the usual way.

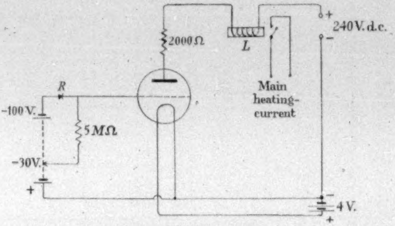


Figure 1. Valve relay circuit.

The disadvantages of the normal form of relay were overcome by using a thermionic valve circuit, figure 1. The regulator contact was placed in the grid circuit of the valve. If the temperature of the bath is too low the contact *R* is not made, and the valve has only a small negative grid bias. The anode current, about 30 mA., is sufficient to operate the relay *L*. As the temperature is slightly raised the contact *R* is closed and the anode current is reduced to less than 1 mA. so that the main heating-current is switched off by the action of the relay *L*. The contacts of this relay consist of mercury contained in two evacuated glass tubes so that sparking is entirely eliminated. This system of regulation has given very satisfactory service, and accurate temperature-control has been effected. At temperatures below 50° C. it has been possible to maintain the temperature constant to within ±0·01° C., while at higher temperatures control to ±0·02° C. could be relied upon.

Temperature-measurement. A mercury-in-glass thermometer was used. Immediately before the commencement of the series of observations it had been calibrated at the National Physical Laboratory for an immersion of 12 cm. The conditions of calibration were accurately observed, and it was therefore unnecessary to apply an emergent-stem correction to the observed readings. The constancy of the temperature of the thermostat was checked by using an uncalibrated platinum thermometer, and was found to be well within the limits previously given.

Time-measurement. A Venner stop-watch was employed for measuring the times of flow of the liquids in the viscometer. Before each observation the watch was fully wound and the differences between successive observations rarely exceeded 0·1 sec., and were usually not greater than 0·05 sec. Any improvement in the method of timing would have been useless without a corresponding increase in the accuracy of temperature-control, and in view of the range of temperature covered the latter would have been exceedingly difficult.

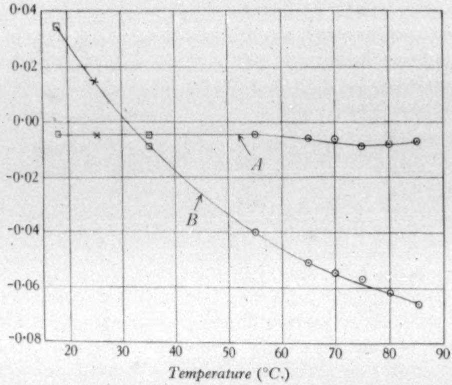


Figure 2. Variation of *A* and *B* with temperature in $\phi/\phi_0 = 1 + A\sqrt{c} + Bc$ for potassium chloride. ⊡ Jones and Talley. × Joy and Wolfenden.

Densities. The densities of water and also of the solutions were obtained from the *International Critical Tables*. Interpolation was necessary in the case of the solutions. The values given in the *International Critical Tables* were used in preference to the results obtained by any one observer as they represent the weighted means of all reliable determinations.

Experimental procedure. The solutions were all prepared by weight and were made as nearly as possible to have round concentrations at 25° C. Since the density of any solution was known at each temperature, the concentration at each temperature could be accurately calculated. The salts were carefully dried in an air oven before using. The water was prepared from a Hartley-type still, and a sample of the water was found to have a conductivity of $1{\cdot}14 \times 10^{-6}$ Ω^{-1}-cm.$^{-1}$ at 18° C. and $3{\cdot}61 \times 10^{-6}$ Ω^{-1}-cm.$^{-1}$ at 75° C.

Between each filling the viscometer and other glassware were cleaned with chromic-sulphuric-acid mixture. No liquid was introduced into the viscometer until the liquid had been filtered.

The temperature was maintained constant for at least 30 min. before an observation was taken. In all cases the mean of at least three consecutive observations was used.

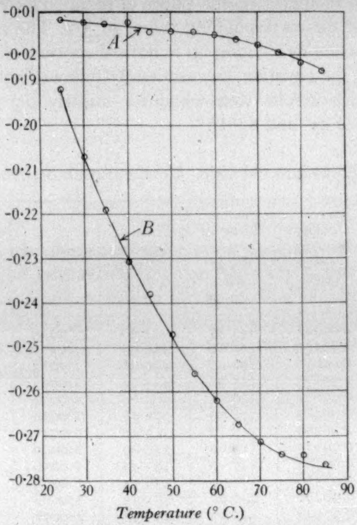


Figure 3. Variation of A and B with temperature in $\phi/\phi_0 = 1 + A\sqrt{c} + Bc$ for potassium sulphate.

§ 3. RESULTS

The results were calculated from the relation

$$\frac{\eta}{\eta_0} = \frac{t}{t_0} \times \frac{d - d_a}{d_0 - d_a},$$

where η is the viscosity of solution at a given temperature,

η_0 the viscosity of water at that temperature,

d the density of solution at the same temperature,

d_0 the density of water at the same temperature, and

d_a the density of air.

To the value so obtained a correction for the difference in the kinetic energies of the two issuing liquids was applied. Detailed results are given in table 1 for potassium-chloride solution of concentration 0·100 moles per litre at 25° C. It will be observed that both the temperature and the concentration vary, but in view of the method employed in calculating the constants no account need be taken of this.

The results for potassium chloride at other concentrations are summarized in table 2, while those for potassium sulphate are given in table 3.

In order to test the validity of the equation

$$\phi/\phi_0 = 1 + A\sqrt{c} + Bc,$$

for each salt at each temperature, the values of $(\phi/\phi_0 - 1)/\sqrt{c}$ at each temperature were plotted against the corresponding values of $\sqrt{c}$. The relation was found to hold for potassium-chloride solution up to normal concentration above 50° C. For potassium sulphate the equation was accurately followed up to the highest concentrations for which results were obtained—namely 0·3 gm.-mol./lit.—at all temperatures between 25° and 85° C.

Table 1. Potassium chloride. Concentration 0·1000*N* at 25° C.

Temperature (° C.)	Concentration of solution (gm. equiv./lit.)	Density of solution (gm./cm³)	Time of flow for water (sec.)	Time of flow for solution (sec.)	$\frac{d-d_a}{d_0-d_a} \times \frac{t}{t_0}$	Kinetic-energy correction	Relative viscosity of solution
18	0·10016	1·00341	645·55	641·48	0·9985	−0·0001	0·9984
25	0·10000	1·00180	548·10	545·55	1·0001	—	1·0001
30	0·09986	1·00038	492·10	490·46	1·0014	—	1·0014
35	0·09969	0·99873	445·70	444·74	1·0025	—	1·0025
40	0·09951	0·99689	406·15	405·64	1·0034	—	1·0034
45	0·09931	0·99487	372·50	372·17	1·0038	—	1·0038
50	0·09909	0·99269	343·70	343·50	1·0041	—	1·0041
55	0·09886	0·99033	318·60	318·69	1·0049	—	1·0049
60	0·09861	0·98783	297·10	297·37	1·0056	—	1·0056
65	0·09834	0·98518	277·90	278·34	1·0063	+0·0001	1·0064
70	0·09807	0·98240	260·80	261·35	1·0068	+0·0001	1·0069
75	0·09777	0·97948	245·67	246·38	1·0076	+0·0001	1·0077
80	0·09748	0·97643	232·40	233·15	1·0080	+0·0002	1·0082
85	0·09715	0·97325	220·70	221·38	1·0081	+0·0002	1·0083

Table 2. Relative viscosity of potassium chloride

Temperature (° C.)	Concentration at 25° C. (gm. equiv./lit.)				
	0·05000	0·1000	0·5000	0·7500	1·000
18	0·9997	0·9984	0·9896	0·9879	0·9837
25	1·0003	1·0001	0·9978	0·9984	0·9994
30	1·0008	1·0014	1·0041	1·0061	1·0098
35	1·0012	1·0025	1·0082	1·0129	1·0188
40	1·0018	1·0034	1·0131	1·0189	1·0273
45	1·0023	1·0038	1·0163	1·0246	1·0349
50	1·0027	1·0041	1·0196	1·0297	1·0406
55	1·0032	1·0049	1·0227	1·0337	1·0469
60	1·0034	1·0056	1·0256	1·0395	1·0528
65	1·0038	1·0064	1·0289	1·0422	1·0577
70	1·0043	1·0069	1·0321	1·0477	1·0627
75	1·0048	1·0077	1·0347	1·0500	1·0675
80	1·0047	1·0082	1·0368	1·0534	1·0717
85	1·0046	1·0083	1·0382	1·0558	1·0747

Table 3. Relative viscosity of potassium sulphate

Temperature (° C.)	Concentration at 25° C. (gm. equiv./lit.)				
	0·01000	0·05000	0·1001	0·2000	0·3000
25	1·0029	1·0126	1·0244	1·0450	1·0685
30	1·0033	1·0131	1·0258	1·0486	1·0736
35	1·0035	1·0136	1·0271	1·0517	1·0779
40	1·0036	1·0141	1·0287	1·0542	1·0816
45	1·0038	1·0150	1·0301	1·0563	1·0850
50	1·0037	1·0160	1·0311	1·0580	1·0878
55	1·0038	1·0166	1·0320	1·0597	1·0909
60	1·0040	1·0168	1·0327	1·0613	1·0934
65	1·0041	1·0172	1·0333	1·0630	1·0953
70	1·0043	1·0175	1·0341	1·0646	1·0976
75	1·0044	1·0181	1·0350	1·0660	1·0989
80	1·0046	1·0187	1·0354	1·0676	1·1002
85	1·0049	1·0187	1·0361	1·0683	1·1022

The values of *A* and *B* at each temperature were calculated from the results in those cases where the Jones-Dole equation had been found to be valid over the whole range of concentrations covered by the method of least squares. The results are summarized in tables 4 and 5. In table 6 the experimental values of the fluidity are compared with the values calculated from the equation with the values of *A* and *B* so determined.

Table 4. Potassium chloride

Temperature (° C.)	Value of constant *B*	Observed value of constant *A*	Theoretical value of constant *A*	Approximate concentration up to which equation is valid
0	—	—	−0·0045	—
18	+0·0330 (3)	−0·0052 (3)	−0·0049	0·15
25	+0·0140 (1)	−0·0052 (1)	−0·0051	—
35	−0·0090 (3)	−0·0051 (3)	−0·0052	0·50
55	−0·0400	−0·0041	−0·0055	1·00
65	−0·0506	−0·0056	−0·0056	Above 1·00
70	−0·0545	−0·0060	−0·0057	
75	−0·0561	−0·0081	−0·0058	
80	−0·0613	−0·0073	−0·0059	
85	−0·0654	−0·0064	−0·0060	
100	—	—	−0·0062	—

The theoretical values of *K* in the equation

$$\frac{\eta}{\eta_0} = 1 + K\sqrt{c}$$

have been calculated by H. Falkenhagen and E. L. Vernon for potassium chloride at 0°, 18°, 25° and 100° C. from the equation

$$K = \frac{1}{60} \frac{\sqrt{\frac{N}{1000}} . eZ}{\sqrt{(DkT)}\sqrt{(8\pi)}} . \frac{1}{\eta_0 u},$$

N where N is Avogadro's number,
e e the electronic charge (e.s.u.),
Z Z the numerical value of the valencies of the ions,
D D the dielectric constant of the solvent,
k k Boltzman's constant,
T T the absolute temperature,
η_0 η_0 the viscosity of solvent, and
u u the mobility of the ions.

Table 5. Potassium sulphate

Temperature (° C.)	Value of constant B	Observed value of constant A	Calculated value of constant A
18	—	—	−0·0128
25	$-0{\cdot}1921_0$	−0·0120	—
30	$-0{\cdot}2071_0$	−0·0122	—
35	$-0{\cdot}2189_9$	−0·0128	—
40	$-0{\cdot}2308_2$	−0·0126	—
45	$-0{\cdot}2379_1$	−0·0145	—
50	$-0{\cdot}2471_7$	−0·0145	—
55	$-0{\cdot}2559_3$	−0·0147	—
60	$-0{\cdot}2619_6$	−0·0154	—
65	$-0{\cdot}2672_1$	−0·0161	—
70	$-0{\cdot}2713_5$	−0·0177	—
75	$-0{\cdot}2740_6$	−0·0193	—
80	$-0{\cdot}2743_0$	−0·0218	—
85	$-0{\cdot}2763_9$	−0·0232	—

Table 6. Potassium sulphate

Temperature (° C.)	Concentration (gm.-mol./lit.)	$\frac{\phi/\phi_0-1}{\sqrt{c}}$	$\frac{\phi}{\phi_0}$ (observed)	$\frac{\phi}{\phi_0}$ (calculated)	Difference (per cent)
25	0·01000	0·029	0·9971	0·9969	0·02
	0·05000	0·056	0·9876	0·9877	0·01
	0·10010	$0{\cdot}075_2$	0·9762	0·9770	0·08
	0·20000	$0{\cdot}096_4$	0·9569	0·9562	0·07
	0·30000	$0{\cdot}117_1$	0·9359	0·9358	0·01
35	0·00997	0·035	0·9965	0·9965	0·00
	0·04985	0·060	0·9866	0·9862	0·04
	0·09979	$0{\cdot}083_6$	0·9736	0·9841	0·05
	0·19938	$0{\cdot}110_2$	0·9508	0·9506	0·02
	0·29890	$0{\cdot}132_2$	0·9277	0·9275	0·02
50	0·00991	0·037	0·9963	0·9961	0·02
	0·04954	$0{\cdot}070_5$	0·9843	0·9845	0·02
	0·09917	$0{\cdot}095_8$	0·9698	0·9709	0·11
	0·19818	$0{\cdot}123_1$	0·9452	0·9456	0·04
	0·29710	$0{\cdot}148_0$	0·9193	0·9187	0·06
75	0·00978	0·044	0·9956	0·9954	0·02
	0·04889	0·081	0·9822	0·9823	0·01
	0·09787	$0{\cdot}108_0$	0·9662	0·9671	0·09
	0·19554	$0{\cdot}140_0$	0·9381	0·9379	0·02
	0·29330	$0{\cdot}166_1$	0·9100	0·9092	0·08

While the actual values of u are unknown at the other temperatures, it was observed that for the four temperatures at which u was known, $\eta_0 u$ varied linearly with the temperature. Hence the values of $\eta_0 u$ at intermediate temperatures may be determined. From these interpolated values, and from the values of the dielectric constant of water determined by Wyman(9), the values of K at the intermediate temperatures may be determined. These are included in table 4.

The small increase in the value of K for potassium chloride indicated by the Debye-Hückel theory was too minute to be detected over the range of temperature covered by previous workers. By considering a wider range of temperatures a definite increase in the value of the constant has been observed with rise of temperature. In view of the difficulties of the work at the higher temperatures the observed value of A cannot be relied upon to more than $\pm 0{\cdot}0015$. The theoretical value is only known to ± 15 per cent, or to about $\pm 0{\cdot}001$ in the case of potassium chloride. The observed and theoretical values are therefore in fairly good agreement.

Unfortunately the theoretical value of A for potassium sulphate is only known at 18° C. The values calculated from the results at 25°, 30° and 35° C. are in good agreement, within the experimental error, with the probable theoretical values at these temperatures. It will be observed that the variation of A with temperature is greater for potassium sulphate than for the chloride.

§ 4. CONCLUSION

The present work gives further confirmation to the predictions of the Debye-Hückel theory; the absolute values of the constant A are in good agreement with the quantitative results of the theory. Moreover, the variation of A with temperature has been shown to be, within the experimental error, that predicted by the theory.

It should be emphasized that this agreement is for one salt only. It seems probable that for salts less simple than potassium chloride the temperature variation will not be that predicted by the theory.

Further, the theory only takes into account one effect of a strong electrolyte on the viscosity of the solvent. Neither the absolute value of the constant B, nor its variation with temperature has yet been determined theoretically.

§ 5. ACKNOWLEDGMENTS

This work was commenced at the suggestion of Prof. H. R. Robinson, F.R.S. I wish to express my thanks for the facilities which he has afforded me at the physical laboratories of the Queen Mary College, and for his interest in its progress.

REFERENCES

(1) Jones and Talley. *J. Amer. chem. Soc.* **55**, 624 (1933).
(2) Jones and Dole. *J. Amer. chem. Soc.* **51**, 2951 (1929).
(3) Joy and Wolfenden. *Proc. roy. Soc.* A, **134**, 413 (1931).
(4) Hood and Hohlfelder. *Physics*, **4**, 6, 208 (1933).
(5) Falkenhagen, H. and Vernon, E. L. *Phil. Mag.* **14**, 538 (1932).
(6) Cox and Wolfenden. *Proc. roy. Soc.* A, **145**, 475 (1934).
(7) Washburn and Williams. *J. Amer. chem. Soc.* **35**, 737 (1913).
(8) Appleby. *J. chem. Soc.* **97**, 2000 (1910).
(9) Wyman. *Phys. Rev.* **35**, 623 (1930).

535.332:546.26:535.245.12

A SPECTROGRAPHIC EXAMINATION OF ARCS BETWEEN PLAIN SOOT CARBONS* AND ITS CONNECTION WITH THE CANDLE-POWER PER AMPERE OF THE POSITIVE CRATER

By Professor J. T. MACGREGOR-MORRIS, M.I.E.E.
and D. E. H. JONES, M.Sc., A.Inst.P.
Queen Mary College, University of London

Received February 23, 1935. *Read in title May* 17, 1935.

ABSTRACT. Analyses of arcs between plain soot carbons of various degrees of purity were carried out with a quartz spectrograph. The normal vertical two-electrode arc and also a horizontal three-electrode arc of *Y* form, with two negative carbons and one positive, were used, parts of the arc flame only being photographed. The object was to find a correlation between the spectra and the candle-power per ampere of the positive crater of the three-electrode arc. The effects on the spectra of the following changes in conditions are given: (i) changing the positive carbon while keeping the negative carbons the same; (ii) special purification of the carbons; (iii) sputtering of the arc; (iv) photographing different parts of the arc (*a*) close to the positive carbon, (*b*) centre of the arc flame, (*c*) close to the negative carbons. Photographic enlargements of the ultra-violet parts of the spectra are given.

§ 1. INTRODUCTION

For many years past an investigation of the photometric behaviour of carbon arcs has been in progress at Queen Mary College with a view to their use as a standard of light. The electrodes of such arcs are arranged in the form of a *Y* with one positive and two negative carbons, the positive being horizontal. An unobstructed view of a perfectly flat, circular positive crater is thus obtained. This arc was originally developed by J. F. Forrest(1) who, using carbons 6 mm. in diameter, obtained the value of 173 cand./mm.² for the crater brilliancy. Later N. A. Allen(2) adapted the arc for larger currents using carbons 10 mm. in diameter, the main modification being that the negatives had to be inclined upwards towards the positive in order to maintain a vertical flat crater. This is the type of three-electrode arc used in the present work, both the angle between the two negative carbons and the angle between the plane of the negative carbons and the positive carbon being adjustable.

A thorough photometric investigation of this type of carbon arc was made by R. C. Fox. He found that only four distinct values of the candle-power per ampere of the positive crater were obtained, the values covering a range of 20 per cent.

---

* Abridgment of a thesis submitted by D. E. H. Jones and approved for the degree of Master of Science in the University of London.

A change in the candle-power per ampere is sometimes brought about by changing the positive carbon and sometimes by changing the negative carbons. Fox made several attempts to correlate the candle-power per ampere with some other property of the carbons. He took photographs of the burning craters and photomicrographs of the craters after burning, and made a cursory examination of the visible part of the spectra of the arc flames. He also determined the rates of carbon-consumption. None of these however could explain the differences although the photographs of the burning craters showed that an increase in the candle-power per ampere was accompanied by a greater homogeneity of the carbon.

It was felt that the spectrographic analyses should be carried further, an excellent quartz spectrograph being available. On account of the great labour involved, however, the present work was confined mainly to the spectroscopy of that part of the arc which lay near to the positive crater, because all the photometric measurements that had been made related to the light emitted from the positive crater.

§ 2. EXPERIMENTAL DETAILS

Spectrographic analyses were made with both the normal two-electrode arc, the carbons being vertical, and the three-electrode arc as developed for a standard of light. The spectra in all cases were photographed with a Hilger quartz spectrograph of type E 3. Only the more important results obtained with the three-electrode type arc are given in the present paper. All other results obtained with both types of arcs confirm those given here. The two-electrode arc was used as the source for photographing the comparison spectra in all cases.

The three-electrode arc-holder consisted of two sheets of uralite, one fixed in a horizontal position on four supporting pillars while the other was hinged and free to swing about one edge of it as axis. The angle between the two could be adjusted to any desired value by means of a semi-circular scale marked in degrees, and a locking thumb-screw. The positive carbon was arranged to slide through a pair of fixed sleeves on the horizontal sheet of uralite, the negative carbons being fixed in similar but moveable holders on the other piece of uralite. These negative carbons could be rotated, so that the angle between them was adjustable.

The two-electrode and the three-electrode arcs were arranged so as to be interchangeable, and an enclosure was formed around them so as to reduce any unsteadiness of the arcs due to draughts. A side image of the arc was thrown on to a screen by means of a lens, and by adjusting the position of the positive carbon so that the image of the end of the carbon coincided with a vertical line in the screen it was ensured that the arc always burned in exactly the same position.

Photography of spectra. Since it was desired to photograph only a part of the arc flame at a time, a vertical slit was placed 3 cm. from the three-electrode arc. This was made from two pieces of sheet iron measuring 1 cm. by 1·8 cm. held in a small holder so that the slit-width was adjustable. The edges of the slit were bevelled and the slit as a whole could be moved across the arc so that any desired portion of the flame could be photographed. For the two-electrode arc the slit was cut horizontally in a small piece of sheet iron.

The analyses were carried out by comparing the carbon spectra with those of R.U. (*raies ultimes*) powder and pure iron. R.U. powder(3) has been developed by the Research Laboratories of the General Electric Company Ltd. and consists of small quantities of about fifty elements incorporated in a base composed of zinc, magnesium and calcium oxides. The quantity of each element has been carefully adjusted so that only the *raies ultimes* and the most important persistent or sensitive lines appear in the spectrum when the powder is burned in an arc. All the elements present are evenly distributed throughout the bulk of the powder so that, under controlled conditions of excitation, an invariable composite spectrum is produced.

The R.U. spectrum was obtained by burning a little of the powder on spectrographically pure Acheson graphite electrodes. Two rods, $\frac{7}{16}$ in. in diameter, were turned down to a diameter of 4 mm. for a distance of 4 or 5 cm. The positive electrode, being the lower, was hollowed out slightly, and about 20 milligrams of the powder were placed in the cavity. The arc-length was adjusted to 6 mm. and the arc was struck by touching with a third graphite rod, the series resistance being such that the current was about 5·5 A.

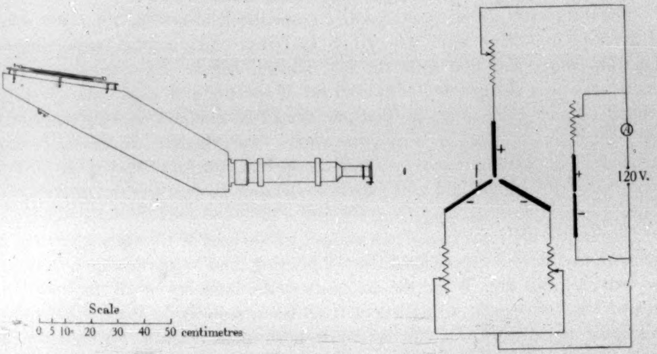


Figure 1.

For the iron spectra Hilger high-purity iron electrodes 5 mm. in diameter were used in the two-electrode arc, the arc current being again 5·5 A.

The three-electrode carbon arc was photographed from the side, a plan of the arrangement being shown in figure 1. In order to reduce exposures to a minimum the arc was placed as near to the spectrograph slit as was consistent with good focusing.

The arc was run off a 120-volt supply, figure 1, and had a series resistance in each negative lead so that the currents through the negative carbons could be adjusted until they were equal. An additional resistance was connected in the positive lead so that variations in the supply voltage could be compensated without disturbing the equality of current through the two negatives. The two-electrode arc was supplied with a separate series resistance.

The arc current for the carbon arc was 12 A. in all cases, the exposure being 60 sec. for the two-electrode type and 10 sec. for the three-electrode type. The angles between the carbons in the three-electrode arc were adjusted to the values given by N. A. Allen(2) for a current of 12 A., i.e. 20° between the positive carbon and the plane of the negatives and 120° separation between the negatives. Under these conditions a vertical, flat and circular positive crater is obtained. The difference in the exposures is due to the fact that two kinds of photographic plates were used; for the two-electrode work Wellington anti-screen plates of speed H and D 450, while for the three-electrode work, Ilford isozenith plates of speed H and D 650 were used. The Ilford plates were backed to reduce halation in the intense parts of the spectra, and no definition was lost as a result of the extra speed of the plates. The corresponding exposures for the iron spectra were 30 sec. and 3 sec. respectively while the spectrum from the R.U. powder was exposed for the time taken by the powder to burn out, which was generally of the order of 1 sec.

The three spectra were photographed immediately above one another by means of a Hartmann diaphragm, the carbon spectrum being in the centre, the iron spectrum above and that from the R.U. powder below. An approximate wave-length scale was also printed with each set of spectra, and several such sets were printed on each plate to ensure that one should be obtained in which the carbon had burned perfectly steadily during the whole of the exposure. In order to facilitate the analysis in the more intense parts of the spectra, one set of spectra in which the exposures of the carbon and iron spectra were much shorter was also taken on each plate.

Despite the shorter exposures necessary in the case of the three-electrode arc, far more trouble was experienced from sputtering than had been experienced with the two-electrode arc. When the arc sputters the intensity of all the lines is increased many times and a number of fresh lines, apparently spark lines, appear; this must be avoided, i.e. the arc must burn quite silently for the whole of the exposure. The effect of sputtering is dealt with more fully later. For the three-electrode arc the crater had to be burned in the correct position, so that it was perfectly flat and perpendicular to the axis of the carbon. The crater was first formed at a slightly higher current, the three carbons were carefully wiped free from adhering particles, and the exposure started immediately on the restriking of the arc. Under these conditions the arc usually burned for 10 sec. without sputtering.

The impurities in the carbons were then determined by comparing the three spectra, i.e. those of carbon, R.U. powder and iron, under a low-power travelling microscope. A set of enlargements of the R.U. and iron spectra, supplied by Adam Hilger, was used for recognizing the lines. The origin of all the *raies ultimes* and sensitive lines is marked on these enlargements so that the elements to which the lines are due can at once be recognized. The wave-lengths of many of the lines in the iron spectrum have also been marked in international angstrom units, so that the wave-length of any unknown line can be found if desired. A visual estimation was made of the intensities of the lines in the carbon spectra at the same time, the lines being divided into four groups as very intense, intense, faint and very faint. From this estimation of intensity the elements were classified as very strong, strong, weak or very weak.

Four different brands of carbons have been used, the various brands being marked as follows:

Witton...G.E.C....Made in England...F...G.E.C.
C. Contradty Noris...Germany
C. Contradty Nürnberg...Marke C
Siemens-Planiawerke A.G. Germany. 308031-A-Homogenkohle-A

These brands of carbons are referred to as *G.E.C. Witton, Noris, Marke C, and Siemens Plania A* respectively.

As has previously been mentioned, spectrograms were taken with both the two-electrode and three-electrode arcs, and in the case of the three-electrode arc they were obtained for the following parts of the arc flame: (i) close to the positive carbon, (ii) at the centre of arc flame, and (iii) close to the negative carbons. This photography of different parts of the arc flame showed that the simplest spectra were obtained by photographing the part of the flame close to the positive carbon, whilst the most complex were obtained by photographing the part close to the negative carbons; moreover it was easier experimentally to obtain spectra of the former portion. Hence, in view of the simplification resulting both in the spectra and in the experimental technique, and more especially because a correlation between spectra and the candle-power per ampere of the positive crater of the three-electrode arc was being sought, most of the photography was done upon the part of the arc close to the positive carbon.

Reading of spectra. Many very faint lines occur in the spectra and these could not be recognized with certainty, particularly when they occur amongst the cyanogen bands or are masked by stronger known lines. This often led to doubt as to the presence of small traces of some elements. These elements, listed below, have been omitted from all the analyses, except when detected with some certainty.

Antimony
Bismuth
Chromium
Cobalt
Germanium
Iridium
Lanthanum
Lead
Lithium
Molybdenum
Osmium
Palladium
Platinum
Rhodium
Ruthenium
Scandium
Silver
Tantalum
Thallium
Yttrium
Zirconium

The barium lines occur amongst the cyanogen bands and great difficulty was consequently experienced in deciding the intensities of these lines. However, barium has been classed as very weak in all cases. Sodium, although it is undoubtedly present in the carbons, was not detected in any of the spectra owing to the fact that the photographic plates were not sensitive to the sodium doublet, λ 5890 and λ 5896, and the other *raies ultimes* of sodium are all extremely faint.

This element, therefore, has been omitted from all the tables. No trace was obtained of zinc in any of the spectra, although some of the *raies ultimes* of this element are very intense.

§ 3. RESULTS

Analyses were carried out by varying the conditions in the following manner. (i) For the part of the arc flame close to the positive carbons: (*a*) changing the positive carbon; (*b*) changing the negative carbon; (*c*) changing the positive and negative carbons; (*d*) special purification of the carbons; (*e*) sputtering of the arc. (ii) For the centre of the flame: (*a*) changing the positive carbon; (*b*) changing the negative carbon; (*c*) changing the positive and negative carbons. (iii) For the part of the flame close to the negative carbons, changing negative carbons. Changes (i) (*a*), (i) (*b*) and (ii) (*b*) were made with both the two-electrode and three-electrode arcs; changes (ii) (*a*) and (ii) (*b*) were made with the two-electrode arc only; and the remaining changes were made with the three-electrode arc only.

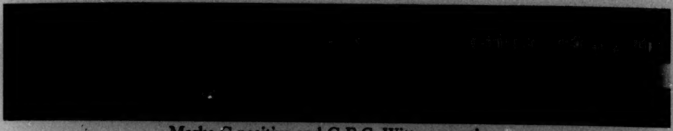
The results in the more important cases are reproduced in full with enlargements of the ultra-violet parts of the spectra. The analyses reproduced were all carried out for the three-electrode arc.

Alteration in the spectrum as a result of changing positive carbons, the candle-power per ampere being changed. The part of the arc flame close to the positive crater was photographed by adjusting the position of the slit in front of the arc so that the edge of the purple image of the flame next to the image of the crater just covered the slit.

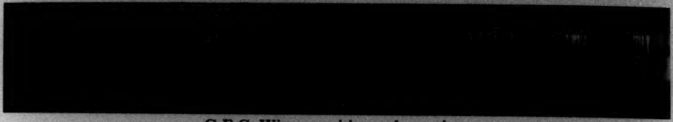
Table 1. Result of changing the positive carbon, the negative carbons being G.E.C. Witton

	Very strong	Strong	Weak	Very weak
Marke C positive, 236 c.p./A.	Carbon, boron, silicon, iron, calcium	Magnesium, aluminium, copper	Manganese, titanium	Arsenic, barium, beryllium, cadmium, nickel, tin, vanadium
G.E.C. Witton positive, 261 c.p./A.	Carbon	Silicon, iron, calcium, magnesium	Boron, aluminium, copper, manganese	Titanium, arsenic, barium, beryllium, cadmium, nickel, tin, vanadium
Noris positive, 278 c.p./A.	Carbon, boron	Silicon	Iron, calcium, magnesium	Aluminium, copper, manganese, titanium, arsenic, barium, beryllium, cadmium, tin, vanadium

The arc investigated in this case was first obtained with G.E.C. Witton carbons as positive and negative. The positive was then replaced by a Marke C carbon and afterwards by a Noris carbon, the first change lowering the candle-power per ampere from 261 to 236 and the second change increasing it to 278. Table 1 gives



Marke C positive and G.E.C. Witton negative



G.E.C. Witton positive and negative

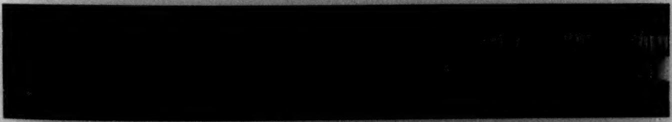
Noris positive and G.E.C. Witton negative

Figure 2. Effect on three-electrode arc spectra of changing the positive carbons.

Commercial Siemens Plania A positive and negative

Specially purified Siemens Plania A positive and negative

Figure 3. Effect on three-electrode arc spectra of purifying the carbons.



Sputtering and sputtering arcs

Silent and sputtering arcs

Figure 4. Three-electrode arc spectra with G.E.C. Witton positive and negative.

the results of the analyses obtained with the three-electrode arc and figure 2 shows the spectra.

The spectra of the arcs that give a high candle-power per ampere are undoubtedly simpler than those that give a lower candle-power per ampere, increasing light-output being correlated with a decrease in the quantities of the impurities present, and particularly of iron, calcium, aluminium, copper and magnesium. This finding has been confirmed in every case by the further analyses which were carried out to find the effect of (*b*) changing the negative carbons but not the positive and (*c*) changing both positive and negative carbons. The change (*b*) was made for both types of arc and for different parts of the three-electrode arc.

Alteration in spectrum due to special purification of the carbons, the candle-power per ampere being constant. Some Siemens Plania A carbons were purified by a remarkably efficient method due to Dr Pirani⁽⁴⁾. The method consists in heating the carbons for about 15 min., in a carbon tube furnace in a stream of chlorine. The temperature is maintained at 2500–3000° C. This purification does not alter the candle-power per ampere of the crater of the arc, the carbons being burned both as positives and as negatives; the value was 261 both before and after purification. The analyses of the spectra of pure and impure carbons was carried out for the three-electrode arc, and the spectra are given in figure 3.

Table 2 shows that the treatment of the commercial carbons has rendered them remarkably pure, for silicon, calcium, arsenic, aluminium, cadmium, manganese and tin have been completely removed. In the case of iron extremely faint traces were visible in the spectrum of the purified carbons but these were so faint and so doubtful that iron is also considered to have been removed by the purification. Silicon and iron were originally classed as strong, calcium as weak, and the other elements as very weak. Magnesium also changes from weak to very weak.

Table 2. Effect of purification of Siemens Plania A carbons used for both poles

	Very strong	Strong	Weak	Very weak
Commercial carbons, 261 c.p./A.	Carbon, boron	Silica, iron	Copper, calcium, magnesium	Arsenic, aluminium, barium, cadmium, manganese, tin, vanadium
Specially purified carbons, 261 c.p./A.	Carbon, boron	—	Copper	Magnesium, arsenic, vanadium

From a consideration of the above facts it appears that the presence of impurities has in itself no effect on the candle-power per ampere of the arc, but that the determining factor is the structure of the carbon matrix. This is unaffected by the purification process although it has been determined by the impurities present during the manufacture of the carbons, and thus originally gave the arcs their various candle-powers per ampere. This view is supported by the fact that the spectra of the arcs having the lower candle-powers per ampere are more complex, and thus show the presence of a greater proportion of impurities, than those having the higher candle-powers per ampere. Again if the operation of the arc leaves the structure of the carbon matrix unaltered, the impurities being volatilized away before the working-temperature of the positive crater is reached, then it is reasonable to expect that chemical purification before use will not alter the candle-power per ampere.

The spectra are shown in figure 3.

Alteration in spectrum due to sputtering of the arc. Sputtering in the carbon arc is primarily due to the flame clearing the loose particles adhering to the tapered sides of the carbons, the particles being thrown out of the arc. It was noticed during visual examination of the spectrum that many fresh lines appeared during sputtering, whilst the intensity of the original lines increased many times. This was particularly noticeable in the ultra-violet region, a fluorescent screen of vaseline smeared on an old photographic plate being used for the visual examination. However, it was felt that some better estimate of the magnitude of the effect would be of service.

The three-electrode arc was used with G.E.C. Witton carbons. The order of photographing the spectra was slightly modified, two sets of three spectra being taken. The first set consisted of the spectrum of the silent arc with spectra of a heavily sputtering arc above and below it. The second set consisted of the spectrum of a heavily sputtering arc with the usual comparison spectra of iron and R.U. powder above and below it respectively. The usual exposure of 10 sec. was given for the carbon spectra, the arc sputtering seven times during an exposure for the spectrum of the sputtering arc. This was arranged by burning the carbons for some time before beginning the exposure and disturbing the arc by slight draughts during the exposure. The spectra are given in figure 4.

Table 3 shows the effect obtained: the quantities of several elements are overestimated as a result of the increase in intensity of their lines caused by sputtering.

Table 3. Effect of sputtering, with G.E.C. Witton positive and negative carbons

	Very strong	Strong	Weak	Very weak
Sputtering arc	Carbon, iron, silicon	Calcium, magnesium, copper	Boron, aluminium, manganese, titanium	Arsenic, barium, beryllium, cadmium, nickel, tin, vanadium, scandium, germanium
Silent arc, 261 c.p./A.	Carbon	Iron, silicon, calcium, magnesium	Copper, boron, aluminium, manganese	Titanium, arsenic, barium, beryllium, cadmium, nickel, tin, vanadium

The total number of lines in the spectrum also increased from about 200 to about 300, practically the whole of the increase being due to iron. No fresh elements were detected with certainty except perhaps scandium and germanium. The following changes are observed: iron and silicon increased from strong to very strong on sputtering; copper increased from weak to strong; titanium increased from very weak to weak. It is thus seen that serious errors may arise as a result of the sputtering of the arc, the amounts present of several impurities being consequently over-estimated. It is therefore essential that the arc shall not sputter but shall burn quite silently during the whole of the exposure if a true arc-spectrum is to be photographed. This rule was strictly adhered to throughout the present work.

Alteration in spectrum along axis of arc. Spectrograms of the three-electrode arc flame were obtained (i) close to the negative carbons, (ii) at the centre of the flame and (iii) close to the negative carbons. The following carbons were used:

	I	II	III
Positive	G.E.C. Witton	G.E.C. Witton	G.E.C. Witton
Negatives	Marke C	G.E.C. Witton	Noris

The spectra from the different parts of the arcs in case II are compared in table 4 and although the positive and negative carbons were of the same brand yet certain changes were observed, more especially between the spectrum obtained close to the negatives and that from the centre of the flame. These changes related to iron, copper, aluminium, manganese and titanium, as will be seen in the table.

Table 4. G.E.C. Witton positive and negative carbons. 261 c.p./A.

	Very strong	Strong	Weak	Very weak
Close to negative carbons	Carbon, iron	Silicon, calcium, magnesium, copper	Aluminium, manganese, boron, titanium	Arsenic, barium, beryllium, cadmium, nickel, tin, vanadium
Centre of flame	Carbon	Iron, silicon, calcium, magnesium, copper, aluminium, manganese	Boron, titanium	Arsenic, barium, beryllium, cadmium, nickel, tin, vanadium
Close to positive carbon	Carbon	Iron, silicon, calcium, magnesium	Copper, aluminium, manganese, boron	Titanium, arsenic, barium, beryllium, cadmium, nickel, tin, vanadium

Visual examination of the spectra shows that they are richer in lines and that the intensities of the lines are greater when they are obtained from the part of the arc flame close to the negative carbons. There is little difference between the spectra obtained close to the positive carbons and at the centre of the flame, the greater change occurring between the centre and the negative carbons. Indeed in the latter case the stronger lines of the spectrum become broadened and shaded on both sides, having an appearance similar to that obtained when the spectrograph is slightly out of focus.

§4. CONCLUSION

The results of this investigation show that undoubtedly a low value for the candle-power per ampere emitted by the positive crater of the three-electrode arc is accompanied, as a general rule, by greater complexity in the spectrum of the arc flame, particularly for the part of the flame close to the positive carbon.

Spectrograms of part of flame close to positive carbon. The most marked change of spectrum with candle-power per ampere of the arc is obtained by changing the positive carbon and not altering the negative carbon. The spectra show appreciable simplification; a steady falling-off in the estimated amounts of iron, calcium, aluminium, copper and magnesium occurs, with increasing value of the candle-power per ampere, both for the two-electrode and for the three-electrode arc.

As would be expected, the differences obtained in the spectra when the negative carbons are changed but the positive carbon is kept the same were not so marked. However, generally similar changes were obtained.

A comparison of the spectra of arcs with different positives and different negatives does not give such large differences as might be expected, only calcium and aluminium showing marked changes. The same general simplification was obtained in the spectrum of an arc giving high candle-power per ampere; with Noris positive and negative carbons the spectrum was appreciably simpler than those of the arcs with G.E.C. Witton or Marke C carbons.

Spectra of carbon arcs yielding the same candle-power per ampere show little differences either in the intensities or in the numbers of lines in the spectra, although the tables of analysis show slight changes in the estimated amounts present of some of the impurities.

The comparison of the spectra of ordinary commercial and specially purified carbons shows that it is possible to remove practically all the impurities without altering the candle-power per ampere, and that therefore the candle-power per ampere is probably determined by the closeness of the structure of the carbon matrix rather than by the chemical properties of the embedded salts. The magnitude and number of the holes these salts leave on volatilization in the arc is the primary deciding factor.

The investigation of the effect of sputtering shows how important it is that the arc should burn quite silently during the whole of the exposure. Appreciable increase in the number and intensities of the lines is caused by sputtering, and leads to over-estimation in the amounts of several impurities present in the carbons.

Spectrograms of the centre of the arc flame. As in the previous section, spectra of arcs yielding different candle-powers per ampere were obtained, firstly by changing the positive carbons, secondly by changing the negative carbons, and thirdly by changing both positive and negative carbons. These experiments lead to the same general conclusions as before, the spectra simplifying with increasing candle-power per ampere though not to an equal extent.

Again the analyses of arcs yielding constant candle-power per ampere show slight changes, though the spectra themselves show little difference as a whole.

Spectrograms of part of arc flame close to negative carbons. Only one series of spectra of this kind was obtained, by varying the negative carbons of the three-electrode arc. Again the results agreed with those obtained from the other parts of the arc flame, although the differences were not so marked as in the case of the part of the flame close to the positive crater.

A comparison of spectra from different parts of the arc flame shows that, for a given arc, they are simplest when obtained from a part of the flame close to the positive crater and most complex when obtained from a part of the flame close to the negative carbons. There is little difference between the spectra of the central part of the flame and the part close to the positive crater.

§ 5. ACKNOWLEDGMENT

The authors would like to express to Prof. A. Fowler, F.R.S., their best thanks for opportunities freely given to discuss their results with him.

REFERENCES

(1) Forrest, J. F. *Electrician*, **71**, 729 and 1007 (1913).
(2) Allen, N. A. *Proc. phys. Soc.* **33**, pt. 2, pp. 62–8 (1921). *Elect. Rev.* Lond., **93**, 238 (1923).
(3) Ryde, J. W. and Jenkins, H. G. *Sensitive Arc Lines of* 50 *Elements* (1930). Adam Hilger, Ltd.
(4) Heyne, G. *Z. angew. Chem.* **45**, 612 (1932).

621.592

LIQUEFACTION OF HYDROGEN BY THE EXPANSION METHOD

By F. SIMON, A. H. COOKE and H. PEARSON,
Clarendon Laboratory, Oxford

Received March 1, 1935. *Read May* 3, 1935

ABSTRACT. Descriptions of some experiments to liquefy hydrogen by the expansion method (*a*) by pure expansion, and (*b*) by the additional use of the Joule-Thomson effect. The yields are of such an order that it seems promising to build large hydrogen-liquefiers on this principle.

Hitherto hydrogen-liquefaction has been exclusively carried out by the Linde process.* As the liquefaction of helium by the expansion method (2, 3) has proved very satisfactory (4), it seemed hopeful to try liquefying hydrogen by this process, and Ahlberg and one of us made a successful preliminary experiment with the apparatus which we had used for helium-liquefaction. In this method the gas to be liquefied is compressed into a container and cooled down as far as possible; for instance, to the temperature of liquid hydrogen for helium liquefaction. Then it is allowed to expand through a valve outside the apparatus, and in doing so it cools according to both the external work done by the gas in the container and the internal work against the van der Waals forces. This method is very efficient for gases with low boiling-point for two reasons: (i) the heat-capacities of the container are very small at low temperatures; (ii) the deviations from the ideal state increase considerably with falling boiling-temperature.

In this paper we describe some investigations of the conditions under which satisfactory results can be obtained for hydrogen. Comparing the liquefaction of helium by this method with that of hydrogen, one has to bear in mind one great difference. In the first case, as we have pointed out already (2), the heat-capacity of the container is negligible, owing to the smallness of the specific heat at the very low initial temperature. This does not hold in liquefying hydrogen, where one has to start at liquid-air temperatures. Hence, in order to keep down the heat-capacity and get a high efficiency one must use a material of a high tensile strength, so as to be able to reduce the weight of the container. However, one must be very careful in selecting the substance of high tensile strength as most materials become very brittle at low temperatures, and in the case of hydrogen there is some danger involved if there is a burst.

---

* See for instance reference (1), p. 315.

De Haas and Hadfield(5) have recently made a very extensive investigation on the properties of a large number of metals and alloys at liquid-hydrogen temperatures, and they found that some alloy steels, especially some with a high nickel-content, do not become brittle even at these low temperatures. Sir R. Hadfield was kind enough to put at our disposal a small cylinder of 144 cm^3 capacity, weighing only 212 gm. It was manufactured from AMF steel (5277 in the paper of de Haas and Hadfield) of a yield-point of 32 ton/in^2 at room-temperature, and 48 ton/in^2 at liquid-hydrogen temperature. Its dimensions were such that using it at a pressure of 150 atmospheres at liquid-air temperature we had a safety factor of 2. The elongation of this steel at the boiling-point of hydrogen had a value of 35 per cent, even larger than that, 32 per cent, at room-temperature.

In the following table we give the heat-capacity of this cylinder as a function of temperature, taking the values of iron and nickel as additive.

T (°K.)	Heat-capacity (cal./°C.)
70	6·9
60	5·0
50	3·3
40	1·8
30	0·8
20	0·2

The heat-capacity of the gas contained in the cylinder at 70° K. and 150 atmospheres is about 12 cal./°C. By means of solid nitrogen it will certainly be possible to start at a temperature between 50° and 55° K. Hence the heat-content of the container will not be too big, especially as the initial amount of gas will then be still higher. But the capacity of the container is not absolutely negligible, so that we have to remember that the following data do not apply for the values one would get with a container with ideal walls; but certainly the difference will not be very great.

For the experiment the cylinder was suspended in a vessel immersed in liquid air, the temperature of which could be reduced by pumping off the evaporated gas. The intermediate space between the vessels was filled with a little hydrogen gas, in order to give heat contact with the external bath. After the cylinder had been filled with compressed hydrogen, and temperature equilibrium had been reached, this exchange gas was pumped away so that the system was thermally isolated. We measured the temperature by means of a copper-constantan thermocouple, one junction of which was fixed to the steel cylinder. The other was connected with a copper block immersed in liquid oxygen, its temperature being measured by means of a platinum resistance thermometer. As the second point for calibrating the thermocouple, we took the boiling-point of hydrogen, and the intermediate values were calculated from the tables given by Giauque(6). The pressure was measured with a Bourdon gauge. In order to avoid big dead volumes, we took a very small instrument fitted with a little mirror and read the deflections with a scale and telescope.

The experiment consisted in letting out successive small amounts of gas by means of a valve situated outside the apparatus. Simultaneous measurements of pressure, temperature and amount of gas were taken. The amount of gas was measured with a gas meter whose readings were sufficiently accurate for our purpose. The amount of hydrogen remaining liquid after expansion to one atmosphere was determined by heating up the apparatus to room-temperature again and measuring the amount of gas liberated. Of course, a correction had to be applied for the amount of gaseous hydrogen inside the apparatus.

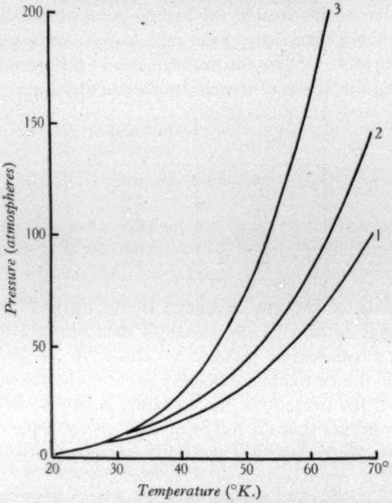


Figure 1. {Pressure, temperature} relation during expansion for three different sets of initial conditions.

Figure 1 gives the connexion between temperature and pressure for three different sets of initial conditions. In the first experiment the vessel remained filled with 1·4 per cent of liquid hydrogen, in the second with 4·4 per cent and in the third 13·8 per cent. By means of the data of figure 1 the values of figure 2, giving the percentage of the volume remaining full of liquid, are derived as a function of pressure at different initial temperatures. We see that for every temperature the three points from our three curves lie on a straight line. The same behaviour was found before for helium over a bigger pressure-range, so that we may extrapolate to higher pressures with some confidence. Thus we can give approximate values for initial conditions to be used in practice; we did not consider it worth while to extend these preliminary experiments to the conditions in question as this would have entailed considerable experimental complications. We can derive that, starting with an initial pressure of 200 atmospheres and a temperature of 52°, which one could easily get by using solid nitrogen (with a vapour-pressure of 6 mm.) a filling-degree between 40 and 45 per cent will be obtained.

One can get a considerable improvement of the yield by adding the Joule-Thomson effect as we suggested in our work on helium⁽⁴⁾. If the gas is not expanded to normal pressure by a valve outside the apparatus but by a valve in thermal contact with the container, then the gas, cooled by the Joule-Thomson effect, will further reduce the temperature of the vessel, if a suitable heat-exchanger is used. In the case of helium we do not generally use this improvement, as the yield is already very high without it, and it complicates the apparatus to a certain extent. Under the conditions prevailing in the case of hydrogen, however, it seems advisable to take advantage of this cooling.

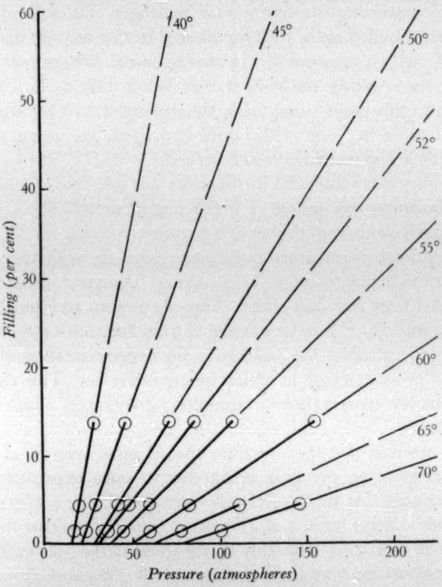


Figure 2. Percentage filling of vessel after expansion from various pressures and temperatures.

We carried out one preliminary experiment of this kind, letting the gas pass through a nozzle which for the sake of simplicity was a fixed one and adjusted so that the whole of the gas went through in about 10 minutes. The gas cooled the cylinder in passing along a spiral soldered on to the outside of it. The conditions of this preliminary experiment were not ideal, as in the early stages while the pressure was still high the gas passed much too rapidly to give a good heat-exchange; hence this experiment only gives a lower limit for the yield.

Actually one will use an initial pressure of about 180 to 200 atmospheres, and a temperature of 50° to 55° K. In our simple apparatus we were able neither to raise the pressure so high as this nor to lower the temperature of the liquid air enough to get conditions similar to those to be used in practice. We therefore cooled the system further by dipping the apparatus into liquid hydrogen for a short time. Then the vessel containing the compressed hydrogen was isolated and the whole apparatus was surrounded again with liquid air. In this way we started our experiment at 45·3° K. with a pressure of 113 atmospheres. We compensated the lower initial pressure by lowering the temperature below that to be used in practice. We shall return to this point later. After the expansion to 1 atmosphere we found the vessel filled to 61 per cent with liquid hydrogen, i.e. 71 per cent of the gas originally present in the container was liquefied.

We will compare this value with the data one can calculate by using the equation of state for hydrogen in this region. For this reason we will consider the following procedure. Directly connected to the high-pressure vessel is a little Linde liquefier. The gas leaving the high-pressure container passes through this and leaves the whole system at the temperature of the container. The yield of the first vessel may then be calculated from our data above. There is now an additional liquefaction in the Linde apparatus which may be calculated from Keesom's enthalpy diagrams(7), which allow us to calculate the yield to a first approximation along the whole temperature and pressure range in which we are interested. This yield is very high on account of the low temperatures in question; it averages, under our conditions, about 50 per cent.

In this way we find that the container should have been filled to 70 per cent with liquid hydrogen, 49 per cent being due to pure expansion. Actually we measured 61 per cent. As we have already mentioned, our preliminary apparatus could give us only a lower limit, and we have to consider too that the heat-capacity of the cylinder was increased noticeably by the spiral of the heat-exchanger soldered to it. In the circumstances we think the agreement good enough.

Calculating the efficiency for the conditions that will be used later on in practice, namely 200 atmospheres and 52° K., one finds about the same yield. Thus one can predict that an apparatus with an efficient heat-exchanger would give a filling degree between 65 and 70 per cent under these conditions.

In view of this high yield, the building of large hydrogen liquefiers seems to be indicated, more especially because of other advantages connected with this principle. The chief point now is to find a material sufficiently reliable and cheap for the container. We are at present engaged on this in the Clarendon Laboratory.

REFERENCES

(1) Meissner, W. *Handb. der Phys.* **11**.
(2) Simon, F. *Z. ges. Kälteindustr.* **39**, 89 (1932).
(3) —— *Proc. roy. Instn*, **22**, 524 (1935).
(4) Simon, F. and Ahlberg, J. E. *Z. Phys.* **81**, 816 (1933).
(5) de Haas, W. J. and Hadfield, Sir R. *Philos. Trans.* A, **232**, 297 (1933).
(6) Giauque, W. F., Buffington, R. M. and Schulze, W. A. *J. Amer. chem. Soc.* **49**, 2343 (1927).
(7) Keesom, W. H. *Leid. Comm.* Suppl. No. 65 *f*.

DISCUSSION

Prof. A. H. Compton asked what pressure of hydrogen was used. Were the walls of an ordinary steel thermos flask strong enough to stand the pressure?

Authors' reply: The pressure to be used in practice will be between 150 and 200 atmospheres. A normal steel dewar flask will not stand these pressures, but when the special steels mentioned are used, the weight can be reduced so far that the cylinder weighs about one kilogram per litre capacity.

551.585.1

AN INVESTIGATION OF THE WET-AND-DRY-BULB HYGROMETER AT LOW TEMPERATURES

By J. H. AWBERY, B.A., B.Sc., F.Inst.P.

AND

EZER GRIFFITHS, D.Sc., F.Inst.P., F.R.S.

Physics Department, National Physical Laboratory,
Teddington, Middlesex

Received February 28, 1935. *Read May* 17, 1935

ABSTRACT. The humidity corresponding to various wet-and-dry-bulb temperatures between −2° C. and −19° C. has been measured. The wet-bulb temperatures were obtained by means of mercury thermometers, thermocouples and resistance thermometers, all three methods agreeing well.

The actual humidity was generally obtained from a specially designed dew-point apparatus, thermojunctions being used to measure the temperature of the metal surface on which the deposit of dew was formed. In addition, tests by the gravimetric method were carried out.

The results are utilized to prepare a table giving the relative humidity at various wet-and-dry-bulb temperatures. These agree, for the most part, with previous tables, but they differ in the region of low dry-bulb temperatures (below about −9° C.) and small wet-bulb depressions (less than 1° C.).

The results are also examined from the point of view of the usual theory which asserts that $(e'-e)=B_0P(\theta-\theta')$, where e' is the saturation vapour pressure at the wet-bulb temperature θ', e the saturation vapour pressure at the dew point (i.e. the actual partial pressure of water in the air), θ the dry-bulb temperature, P the barometric pressure and B_0 the psychrometric constant. Although above 0° C. this formula is found to hold well, yet below 0° C. it is found that B_0 varies in a complicated manner with the temperatures. It tends to infinity when the wet-bulb depression tends to zero, and there is a further variation superimposed on this one; at a constant value of the depression, B_0 passes through a minimum and then through a maximum as the dry-bulb temperature decreases.

§ 1. INTRODUCTION

THE question of the determination of the moisture content of the atmosphere is one which has interested investigators for the past 100 years and yet comparatively little work has been done on the subject outside the range of normal atmospheric temperatures. In recent years the problem of the measurement of humidity in cold stores has come to the fore on account of the increasing attention which is being given to the scientific study of the storage of foodstuffs by refrigeration.

Here the investigator is confronted with a two-fold problem, (1) the adaptation of the known methods of measuring humidity to temperatures below the freezing-point of water, and (2) the calibration of such apparatus in the temperature range in which he is interested.

The present paper deals with the measurement of humidity in the range from 0° to −20° C. particularly with reference to the wet-and-dry-bulb type of hygrometer and the tables to be used with it when one of the thermometer bulbs is coated with ice. There are inherent difficulties associated with hygrometric measurements at low temperatures. For example, the weight of water vapour per unit volume of air is relatively small; the observer has to manipulate the apparatus without disturbing the humidity of the air: accurate data as to the saturation moisture content of air at low temperature are lacking.

Consequently with the hygrometric instruments available it is not possible to obtain the humidity of the atmosphere with a high degree of precision when working at temperatures well below the freezing-point of water.

§ 2. TYPES OF HYGROMETER

Among the many forms of hygrometer which have been devised from time to time, only four are in common use at ordinary temperatures—the gravimetric, dew-point, wet-and-dry-bulb, and the extension-of-hair types.

The first two hygrometers depend on well-founded physical principles and give readings that can be converted to values of relative humidity without prior calibration. As regards the wet-and-dry-bulb instrument, it is conceivable that in the future a complete theory will be worked out, but the resulting formula will probably contain one or more constants to be determined empirically.

The present position is that the theory of the wet-and-dry-bulb hygrometer, or psychrometer, is incomplete, and it is necessary to prepare tables for any ranges of temperature and humidity over which it is to be used.

In principle the hair hygrometer is the simplest of all forms of hygrometer but it is necessary to calibrate each instrument before use.

§ 3. INSTRUMENTS FOR ROUTINE USE AT LOW TEMPERATURE

The gravimetric method is unsuitable for routine use even at ordinary room temperatures, owing to the length of time occupied in obtaining a reading, but it is particularly difficult at low temperatures such as are met with in the cold storage of foodstuffs, owing to the very small amount of water which the air contains, even at high relative humidities. For example, if the temperature is −5° C., it is necessary to pass 293 litres of air through the absorption tubes before even 1 gm. of water can be collected, as compared with 58 litres of air necessary at 20° C.

Moreover, to weigh 1 gm. of water to an accuracy of 1 per cent (i.e. 0·01 gm.) is not easy when the temperature is −5° C. and the total weight dealt with—U-tube and desiccating agent—is 20 gm. or more.

When relative humidity, rather than moisture content, is required, the greatest obstacle is not the experimental difficulty, but the fact that the moisture content for saturation has not been directly measured at low temperatures, and an uncertainty is therefore introduced owing to the extrapolation of an empirical formula.

Turning to the dew-point method in its application to low temperature work, we find that the first difficulty is that of cooling the silver surface below the temperature of the surroundings. Volatile liquids such as ether do not evaporate readily at low temperatures, and are thus less effective than at normal temperatures; further, the vapours from them are often deleterious to the foodstuffs which may be preserved in the cold store. A second and more important point is that owing to the increased viscosity of the cooling fluid, the heat transfer between the fluid and the metal mirror is slowed down and it may no longer suffice to measure the liquid temperature as an indication of the surface temperature of the metal. Also the dew is liable to deposit in thick clusters unless particular care is taken to ensure cleanliness of the surface. The dew-point method cannot be operated in a very small enclosure since owing to the small moisture content of the air, the quantity of water removed to give a visible deposit may be sufficient to disturb appreciably the relative humidity of the remaining air. The small moisture content also adds to the necessary equipment, in that it becomes of importance to provide some means of keeping the air in gentle movement. Finally, it must be remembered that with the standard forms of dew-point apparatus, the observer must be in close proximity to the silver surface to note the point at which dew first appears. It is usually undesirable to open up a cold store too frequently (as regards the holds of a refrigerated ship it would be impossible). Yet from many points of view, the dew-point instrument is a desirable form, and it is possible that most of the obstacles to its use at low temperatures could be avoided by an instrument using a photo-electric cell to observe the dew, and specially designed in other respects to be adapted for low temperature work. It is hoped to follow up this subject, and to prepare a note later dealing with such an instrument. Meanwhile, for routine work the choice lies between the wet-and-dry-bulb type and the extension of hair type. The latter has a considerable time lag and is known to be subject to sudden changes of calibration, especially if subjected to very dry atmospheres. Nevertheless it is possible that it would be a highly useful instrument if the conditions which bring about these sudden changes were fully known, and if the lag could be shortened. By avoiding the undesirable conditions (or recalibrating immediately after they had been encountered) the instrument could then be regarded as reliable. This type of instrument is at present under investigation, and the results will be made available later.

The remaining instrument, the psychrometer, appears on balance to have less disadvantages than other types, and is in fact fairly frequently used in practice. Its chief disadvantage is that at low temperatures the wet-bulb depression becomes very small, so that a slight error in reading the wet-bulb temperature has a considerable effect on the relative humidity. Thus at −10° C. an error of 0·2° C. in the wet-bulb reading involves on the average about 6 per cent error in the relative humidity, whereas at +20° C. the same wet-bulb error only leads to an average error of 1 per cent relative humidity. The instrument suffers also from the defect that it requires attention from time to time to ensure that the wet bulb is still glazed with ice, though less frequently when the bulb is ice-covered than at higher temperatures. Nevertheless the wet-and-dry-bulb instrument is in our experience the most promising of those available for routine work, especially in view of the improvements in temperature measurement made possible in recent years. In view of this, and considering also that the published tables are believed to be based on formulae deduced from experiments at higher temperatures, it was decided to make a study of the wet-and-dry-bulb instrument at temperatures down to −20° C.

§4. DESCRIPTION OF APPARATUS

For this investigation it was necessary so to arrange the apparatus that the observations could be taken without necessitating the presence of the observer in the same room as the hygrometers; otherwise the water vapour given off in respiration would be a serious disturbance.

A metal box measuring 7 ft. × 6 ft. × 7 ft. was constructed and placed in the centre of a chamber measuring 13 ft. × 16 ft. × 14 ft. This chamber* was provided on floor, ceiling, and four walls with piping for the circulation of cold brine, by which any temperature down to −20° C. could be maintained in the atmosphere outside the metal box. Fans were installed to promote uniformity of temperature in the interspace. The various psychrometers under study, together with the standardizing instruments, were installed near the windows of the metal enclosure. The two glass windows in two adjacent sides permitted of observations being made on the dew-point surface and on the mercury thermometers.

The air in the metal enclosure was maintained in continuous circulation by the fan and duct system shown in figure 1. The air was delivered through ports in one side and abstracted through ports in the opposite side. Direct flow between the two sets of ports was prevented by the baffle plates shown. Control over the humidity of the circulating air was obtained by the use of calcium chloride or other desiccating agent which was placed on trays in the ducts.

Various forms of the wet-and-dry-bulb type of hygrometer were installed in the metal enclosure together with a silver thimble and a disc form of dew-point apparatus. The formation of dew was observed visually through the window of the metal enclosure by an observer in the interspace. When electrical methods of temperature measurement of the dew-point surface were in use, this observer signalled the appearance and disappearance of dew to another observer outside the chamber, where the potentiometer was situated.

A pipe leading through the wall of the enclosure enabled a sample of air to be led into U-tubes containing pumice soaked with concentrated sulphuric acid for use in the gravimetric method. The balance for weighing the U-tubes was in the chamber but outside the enclosure, so that the tubes could be weighed without subjecting them to any changes of temperature.

---

* The chamber used was one at the Ditton Laboratory, East Malling.

§ 5. WET-AND-DRY-BULB INSTRUMENTS

The pair of observations which are given by a psychrometer can either be obtained by measuring the dry-bulb temperature and the actual wet-bulb temperature, or else by measuring the dry-bulb temperature and the depression of the wet bulb. The latter method has distinct advantages, for the actual temperatures need not be known so accurately if the depression is measured directly. For example, a depression of 1° C. at a dry-bulb temperature of −9° C. corresponds to 63 per cent relative humidity, and the same depression at −10° C. corresponds to 61 per cent.

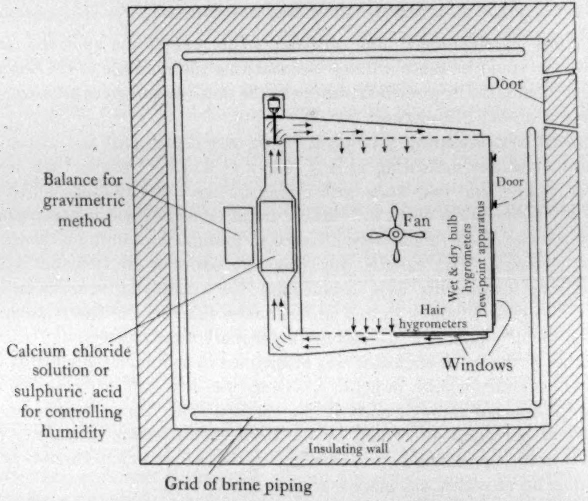

Door
Balance for gravimetric method
Door
Fan
Wet & dry bulb hygrometers
Dew-point apparatus
Hair hygrometers
Calcium chloride solution or sulphuric acid for controlling humidity
Windows
Insulating wall
Grid of brine piping

Figure 1. Plan of metal enclosure and chamber.

There is thus a considerable advantage in using thermocouples or resistance thermometers arranged differentially to give the depression directly, and such a procedure has the additional advantage that temperature differences can be read to less than 1/10° C., a quantity which, although adequate for 1 per cent accuracy in humidity at ordinary room temperatures, is not so at temperatures well below 0° C.

In view of the fact that the wet-bulb depression might conceivably vary with the size of the thermometer bulb, it was necessary to take observations with the more usual mercury in glass type of thermometers, as a check on the other forms of wet-and-dry-bulb hygrometer. Actually the electrical and the mercury thermometers were found to agree within the limits of error of the latter.

The thermocouple psychrometer was made of copper and constantan wires each of 26 gauge, four junctions being used in series. The wet junctions were covered with about 2 in. of linen sewn into the form of a tightly fitting sheath after placing into position. Both junctions were placed in a 2-in. tube inserted in the wall of the enclosure, through which the air was drawn by means of a fan. The wet junction was downstream, so that it did not affect the air passing over the dry bulb. The electrical measurements were made on a potentiometer outside the chamber.

The resistance thermometers were of platinum wound on a mica cross, and enclosed in a glass sheath of length about 12 in. and diameter 8 mm. The actual thermometer coil was 8 cm. long, and in the case of the wet bulb, a linen sheath 13 cm. long was applied outside the glass sheath.

The two thermometers were inserted in a glass tube with a mica partition down the centre to separate them, so as to ensure that the wet-bulb thermometer should not affect the readings of the other.

The mercury thermometers were of the solid stem type having a cylindrical bulb about 2 cm. long and 6 mm. in diameter, the customary form of linen sleeve being fitted to the wet bulb. They were of range -20° to $+5^\circ$ C. divided to $0{\cdot}1^\circ$ C., each division occupying approximately 1 mm. These thermometers were mounted in the standard form of Assmann casing with an electrically operated fan. This instrument was suspended from the roof of the enclosure, near one of the windows, through which the observations were taken. When it became necessary to re-moisten the wet bulb, the whole instrument could be withdrawn through a hole in the roof. On these occasions care was taken that the layer of ice remaining on the wet bulb was melted before the instrument was replaced in position for use.

§ 6. STANDARDIZING INSTRUMENTS

In order to correlate the readings of the psychrometer with true humidities, the latter must be measured with instruments for which the theoretical formulae contain no adjustable constants. As was pointed out in the first paragraph, the gravimetric and the dew-point methods are the only ones fulfilling these conditions.

The difficulties of each of these have been referred to earlier, and it has been in fact necessary to modify the dew-point instrument for the present investigation.

Initially the Regnault thimble type was used, cooling being effected by passing paraffin, previously cooled with solid carbon dioxide, through the silver thimble. The results failed to agree with those of the gravimetric method, and it was suspected that the thermometer in the liquid contained in the thimble measured a temperature different from that of the silver surface on which the dew was deposited. This was verified by attaching a thermocouple to the silver surface itself, this experiment showing that there was in fact a temperature discontinuity across the interface, which increased with the dew-point depression.

A dew-point instrument in which the actual metal temperature is observed thermoelectrically is more convenient to construct if the metal mirror takes the form of a disc, and in addition, it was found that observation of the dew through the window of the enclosure could be made more conveniently when the deposit occurred on a flat surface. Consequently, the dew-point apparatus used in the main body of the work was constructed of a plane disc about 6 cm. in diameter, soldered to a brass block, hollowed out to form channels along which the cooling fluid could flow in contact with the disc (see figure 2). The channels were designed with a view to obtaining uniformity of temperature over the surface. The cooling fluid, as with the earlier type, was pre-cooled paraffin. The temperature of the silver was observed by means of a copper constantan thermocouple of 36 s.w.g., with its junction pressed against the rear surface of the disc by an ebonite plug. This form has been found very satisfactory in use, and the values of the humidity found by means of it agree substantially with those found by the gravimetric method.

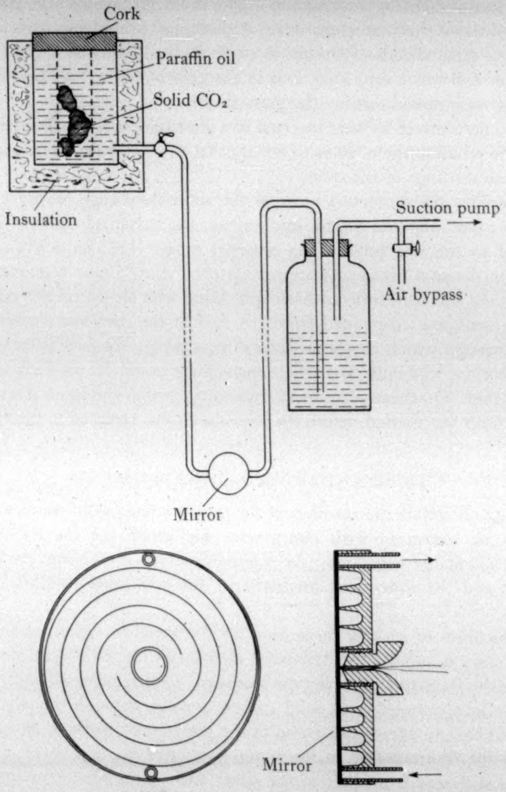


Figure 2. Dew-point apparatus.

The last-mentioned method suffers from the practical disadvantage that as the temperature is lowered the absolute weight of water per unit volume for a given relative humidity decreases and also there is some doubt as to the moisture content of saturated air under these conditions. It was therefore only used in this investigation as a check to ensure that the main standard, the dew-point apparatus, was functioning satisfactorily. Three large bore U-tubes were employed, in series, and it was generally found that the last U-tube of the series increased in weight by an insignificant amount. Errors such as those due to absorption on the glass surface, were minimized by the use, during weighing operations, of a counterpoise of similar tubes suspended from the opposite arm of the balance.

The absorbing medium was prepared by heating broken pumice to bright redness, and dropping it into concentrated sulphuric acid.

By means of a small motor-driven pump, the air was drawn out of the enclosure through the U-tubes, and then through a dry gas meter situated outside the cold chamber. In a typical experiment, volumes of the order of 40 cubic feet of air were used, the resulting increase in weight of the U-tubes due to the absorption of the water vapour being of the order of 2 gm.

§ 7. RESULTS

In the main experiments, which consisted of a series of observations of the wet-and-dry-bulb readings over a range of humidities and a range of temperatures between 0° and $-20°$ C., the standard hygrometer directly utilized was the dew-point instrument. It was found difficult to operate the gravimetric method so as to obtain frequent readings, and moreover, as has been previously mentioned, it suffers from the defect that at very low temperatures, the moisture content of air at saturation is unknown. It may, however, be regarded as having indirectly provided a second standard in these experiments, because a number of comparisons were made between it and the dew-point instrument used in the main work. Further, there are a few of the main experiments in which the gravimetric and not the dew-point method was used as standard.

When using the gravimetric instrument, the air was metered in a dry condition at the temperature prevailing outside the chamber; the difference between its pressure at the meter and the barometric pressure was observed, and the volume which the same mass of dry air would occupy at the barometric pressure and the temperature of the enclosure in the cold chamber was calculated by the law $pv/T = \text{constant}$, where p is the pressure, v the volume, and T the absolute temperature. The mass of water collected in the sulphuric acid tube, divided by the corrected volume, gives directly the moisture content of the air in the enclosure, in grams per cubic metre, since the volume of the air when dry is the same as that of the moist air, by Dalton's law.

To estimate the relative humidity, the moisture content thus calculated must be divided by the moisture content of saturated air. This was obtained from the tables published by the Prussian Meteorological Institute(1) which give a table of the difference between the moisture content f and the vapour pressure e, calculated by the formula $f=289{\cdot}4\ C/T$, where T is absolute temperature. We are not aware of any experiments to test the truth of this formula at low temperatures, and it might be mentioned that experiments at temperatures between 40° and 100° C. show that it is not quite accurate in this region(2).

The relative humidity, when the wet-bulb temperature is below 0° C., can be defined in two ways, according as saturation is taken to refer to the condition which would be in equilibrium with ice or with super-cooled water at the dry-bulb temperature. In this calculation, and indeed throughout the paper, we have taken the latter standard of reference. This seems to us to be decidedly preferable on many grounds. In the first place, at moderately low temperatures, super-cooled water can actually be obtained. Consequently, air in contact with it would take up sufficient water to become practically saturated in this condition. If the same air was transferred to an enclosure where ice particles were exposed, it would deposit some of its moisture. That is, relative to the ice standard, the relative humidity was originally above 100 per cent without any tendency to deposition. Again, by using the ice condition as standard of reference, a discontinuity in relative humidity is artificially introduced when the dry bulb alters from slightly above to slightly below 0° C. On these and similar grounds, we prefer to use the state of saturation with reference to super-cooled water, as our standard of reference. This agrees with the practice of the British and Prussian Meteorological Offices, but not with that of the U.S. Weather Bureau.

The dew-point observations were reduced with the aid of the vapour pressure tables given in International Critical Tables, using the values of vapour pressure over ice at the dew-point temperature to obtain the actual vapour pressure, and dividing this, as explained above, by the vapour pressure over water at the dry-bulb temperature; the degree of agreement between the two methods is shown in table 1 below in which some of the entries are the means of two determinations.

Table 1. Comparison of gravimetric with dew-point hygrometer

Dry bulb (°C.)	Wet bulb (°C.)	Relative humidity (per cent)	
		Dew-point method	Gravimetric method
− 5·8	− 6·2	84	83
− 9·5	− 10·8	42	45
− 10·0	− 11·2	45	47
− 12·1	− 13·3	43	47
− 14·0	− 14·6	56	60
− 16·3	− 17·0	42	42
− 16·4	− 17·2	40	46

The tendency is for the gravimetric method to give the higher result, but the mean difference between the instruments is only 3 per cent in relative humidity. This may be due to the uncertainty in the saturation moisture content. On the other hand, it might also indicate that the dew point gives low relative humidities, as would happen if the apparent dew-point temperature was below the true. The error in the temperature would have to be about 0·6° C. to account for the whole of the difference, and it is unlikely that the observation could be in error by this amount, particularly in view of the agreement found between the appearance and disappearance of the dew. It may therefore be concluded that in general the dew point should give the humidities correctly, at any rate to 2 per cent. This degree of accuracy seems to be all that is required, since repetitions of the observations at a given dew point show variations in the wet- and dry-bulb observations which correspond to 2 per cent or more in the relative humidity. In the main series of experiments, the air in the enclosure was first brought to the desired humidity and temperature, and observations of the wet-and-dry-bulb thermometers taken at intervals, alternately with those of the dew-point instruments. The mercury thermometers for the wet- and dry-bulb readings were read by an observer situated in the interspace between the metal enclosure and the cold chamber, and he also controlled the cooling of the dew-point instrument, and signalled the appearance or disappearance of dew to a second observer outside the chamber. The latter also took readings of the thermoelectric and resistance thermometer psychrometers.

It was invariably found that the three forms of psychrometer agreed with each other within their limits of error. The dry-bulb temperature was taken to be that shown by the mercury thermometer, whilst the wet-bulb temperature was taken as the mean of those given by all three instruments. In a typical experiment, the conditions were maintained approximately constant for a period of about 2 hours during which time about a dozen observations of each instrument were obtained. The various wet-bulb readings, dry-bulb readings, and dew-point observations were averaged, and the true relative humidity calculated from the latter.

The results of all the experiments carried out are shown in table 2.

§ 8. DISCUSSION OF RESULTS

Attention may be drawn to one feature of the low-temperature data given in table 2, namely, the rapid alteration of relative humidity with change of the wet-bulb temperature. For example, with a dry bulb of $-3·5°$ C. an alteration of only 0·7° C. in the wet bulb corresponds to a 7 per cent alteration of relative humidity (from 34 to 21 per cent). Still more striking are the observations at $-10·7°$ and $-10·8°$ C. dry bulb; in this case a change of 1° C. in the wet bulb is associated with an alteration from the relatively moist condition of 77 per cent to the relatively dry one of 51 per cent relative humidity.

To express the results concisely, it is more convenient to draw up a table (or graph) of relative humidity as a function of wet-bulb depression for successive values of the dry-bulb reading, rather than to take the wet- and dry-bulb readings themselves as variables.

The usual psychrometric formula, valid for temperatures above 0° C., is $(e'-e)=B_0P(\theta-\theta')$, where e' is the vapour pressure at the wet-bulb temperature

Table 2. Observed humidities corresponding to various wet- and dry-bulb temperatures

Temperature (° C.)		Relative humidity (%)	Temperature (° C.)		Relative humidity (%)
Dry bulb	Wet bulb		Dry bulb	Wet bulb	
−2·23	−5·74	30	−9·42	−11·51	30
−2·43	−5·87	32	−9·47	−10·84	45*, 43
−2·65	−5·78	39	−9·54	−10·73	46*, 41
−3·10	−6·80	22	−9·60	−11·68	27
−3·15	−7·14	22	−9·62	−11·74	29
−3·27	−6·77	31	−9·67	−10·96	51
−3·45	−6·77	29	−9·96	−11·22	45, 47*
−3·53	−7·24	21	−10·39	−11·64	48
−3·55	−6·51	34	−10·47	−12·12	36
−3·55	−6·89	27	−10·70	−11·94	51
−3·55	−7·19	21	−10·80	−10·90	77
−4·05	−7·64	21	−10·93	−10·88	86
−5·07	−5·62	87	−11·02	−11·33	66
−5·11	−5·60	82	−11·37	−11·19	77
−5·12	−5·59	83	−11·57	−12·67	49
−5·19	−6·70	63	−11·80	−13·08	49*
−5·44	−6·35	73	−11·80	−13·16	43
−5·64	−7·30	53	−11·96	−13·25	43
−5·74	−6·19	86	−12·02	−13·26	42
−5·74	−6·19	87	−12·16	−13·07	51
−5·76	−6·19	84	−12·26	−13·43	44
−5·82	−6·17	84	−12·37	−13·44	45*, 42
−5·85	−6·27	83*	−12·39	−13·75	42
−5·88	−9·03	21	−12·65	−14·12	31
−5·92	−7·92	42	−12·68	−13·98	37
−6·09	−9·22	22	−12·68	−14·23	22
−6·21	−6·61	82	−12·84	−14·40	29
−6·34	−9·47	23	−13·63	−13·78	67
−6·45	−8·64	39	−13·79	−13·99	69
−6·51	−9·33	26	−13·99	−14·57	60*, 56
−6·54	−8·13	52	−14·45	−15·75	30
−6·59	−8·38	46	−14·56	−15·90	28
−6·74	−9·27	31	−14·68	−15·87	31
−6·84	−9·42	29	−15·67	−16·43	43
−6·91	−9·74	20	−16·05	−16·74	50
−7·20	−9·32	42	−16·06	−16·97	38, 48*
−7·24	−7·40	86	−16·10	−17·04	44
−7·26	−7·66	82	−16·28	−17·30	27
−7·38	−7·70	81	−16·29	−17·05	42, 42*
−7·44	−7·59	87	−16·63	−17·82	27
−7·51	−7·64	91	−16·66	−17·53	41, 48·5*
−7·52	−7·63	87	−16·75	−17·51	45*
−8·10	−9·81	46	−16·76	−17·93	28
−8·22	−9·82	43	−17·03	−17·70	51*
−8·41	−8·52	86	−17·36	−18·18	47*
−8·58	−8·66	87	−17·72	−18·72	34
−8·73	−8·94	78	−17·82	−18·22	61
−8·88	−11·33	22	−17·92	−19·01	32
−8·92	−10·12	51	−18·22	−18·80	45

* These relative humidities were obtained by the gravimetric method.

θ', e is the vapour pressure at the dew point, θ is the dry-bulb temperature, P the barometric pressure, and B_0 is the psychrometric constant. There is evidence that above 0° C. the "constant" B_0 is a linear function of the wet-bulb temperature.

A preliminary examination of the data obtained in the present investigation showed that B_0P is not a simple function of the above variables at low temperatures. Indeed it is evident that below 0° C. B_0 cannot be either constant or a function of wet-bulb temperature only, since with ice-covered bulbs it is possible to have $\theta=\theta'$ without $e=e'$ since the saturation vapour pressure over ice is not the same as that over an undercooled water surface. When $\theta=\theta'$ and $e=e'$, B_0 is infinite. This fact by itself would suggest that B_0 might be a function of $(\theta-\theta')$, the wet-bulb depression, which took indefinitely large values as $(\theta-\theta')$ approached zero. The observations do show that B_0 increases rapidly as the wet-bulb depression tends to zero, but it is not a simple function of the latter, since at a fixed depression it is found that there is a considerable range of values of B_0. Consequently if B_0 is introduced into the theory at all, it must be taken as depending on two quantities, wet-bulb depression and dry-bulb temperature. Since two independent variables must be taken into account for the formula it seems simpler to deal directly with the relative humidities than to introduce the parameter B_0.

No attempt was therefore made to deduce an equation, and the observations were smoothed graphically. As a first step, the relative humidities were plotted against wet-bulb depression, a separate curve being drawn for each 1° range of dry bulbs. These curves were for the most part spaced about 4 per cent of relative humidity from each other, and the extreme separation was only 7 per cent.

With the aid of this preliminary chart, a correction was applied to each observation so that it gave the relative humidity corresponding to the observed depression, at the nearest integral dry-bulb temperature. Thus, the first observation in table 2, which gives the relative humidity corresponding to a dry bulb of −2·23° C. and a depression of 3·51° C. was corrected to give the relative humidity at a dry-bulb temperature of −2·0° C. and a depression of −3·51° C. Similarly the second entry of the table was corrected to a dry bulb of −2·0° C. and the third to −3·0° C. In no case did the correction exceed 3 per cent relative humidity, and it is probable that any error it introduced would always be less than ½ per cent.

The corrected results were again plotted against wet-bulb depression, and smoothed values read off at round values (0·0°, 0·2°, 0·4° ... 4·0° C.) of the depression, to give a table (not reproduced) which we call table A.

At this stage it is known that the results for any dry bulb vary regularly with depression, and represent the observations which were taken at about that dry-bulb reading. Individual curves could, however, lie slightly low or slightly high as compared with the others, on account of experimental error, and it is therefore necessary to plot the results against dry-bulb temperature, in this case using one curve for each fixed depression. Smooth curves were drawn through these points, and values again read off at the integral dry-bulb temperatures, and for depressions of 0°, 0·2° ... 4·0° C. to give table B. This process disturbs the smoothness of the curves of humidity against wet-bulb depression, and these were therefore re-drawn from the last table. However, it is important that in this final smoothing of the results from table B, the final curves should not depart more than can be avoided from those of table A, which represents the observations more directly. Consequently the results of table A were also plotted on the same sheet, and used as a guide in drawing the lines to represent the points. The net result is to obtain a able in which the relative humidity varies smoothly when considered either as a function of wet-bulb depression or of dry-bulb temperature. The final results are shown in table 3. The fidelity with which this table represents the observations is shown in figure 3. To prepare this, the humidity corresponding to the observed wet- and dry-bulb temperature was obtained from table 3 for each experiment individually, and is plotted against the humidity observed by the dew-point or gravimetric method in the experiment concerned.

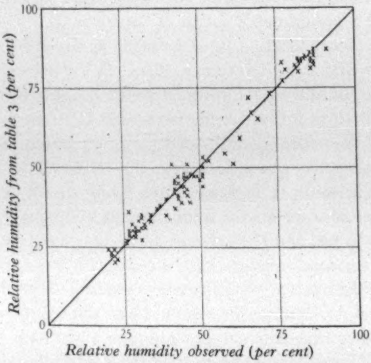


Figure 3. Comparison of tabular values with observed values.

From this it will be seen that in general the values read from the table are within about 4 per cent of the observed values; closer agreement cannot be expected, since the experimental error in an individual experiment is probably of this order of magnitude (as may be noted, for example, by comparing the fourth, sixth and seventh entries in table 2). Further evidence that the errors are distributed at random is provided by figure 4, which shows the frequency distribution of the various errors.

Nevertheless, it is desirable to ensure that no part of the discrepancies is due to the table running consistently high in one area and low in another. (Figure 3 shows that if such areas exist, they occur in pairs which balance in this way, since otherwise there would be a preponderance of points either above or below the line.) To this end the errors (i.e. the differences between humidities observed and those read from table 3) were grouped into a table of double entry, with dry-bulb and wet-bulb temperatures as variables. The purely random errors in such a table tend to obscure the general trend, so that ranges of one or two degrees in the temperatures

Table 3

Wet-bulb depression (° C.)	Dry-bulb temperature (°C.)																			
	0	−1	−2	−3	−4	−5	−6	−7	−8	−9	−10	−11	−12	−13	−14	−15	−16	−17	−18	−19
−0·1	—	100	100	99	98	97	95	93	90	87	83	80	77	74	71	69	67	65	64	62
0	100	99	98	98	97	96	94	92	89	85	82	78	75	73	69	67	66	64	62	60
0·2	96	95	94	93	92	91	89	86	83	80	76	73	70	67	64	62	60	58	56	54
0·4	92	91	90	89	88	86	84	81	78	74	71	68	64	62	58	56	54	52	50	48
0·6	88	87	86	85	83	81	78	75	72	69	66	63	59	56	53	51	48	46	44	42
0·8	84	83	82	80	79	77	73	70	67	63	60	57	54	51	48	45	43	40	39	36
1·0	80	79	78	76	74	72	68	65	62	58	55	52	49	46	42	40	37	35	33	30
1·2	76	75	74	72	70	67	64	60	56	53	50	46	43	40	37	34	32	29	27	24
1·4	73	71	70	68	65	63	59	55	51	48	44	41	38	35	32	29	27	24	22	
1·6	69	68	66	64	61	58	54	50	46	43	39	36	33	30	27	24				
1·8	65	64	62	60	57	53	49	46	41	38	35	31	28	25	22					
2·0	62	60	58	56	53	49	45	41	36	34	30	26	24							
2·2	58	56	54	52	49	45	41	37	32	29	25	22								
2·4	55	53	50	48	44	40	36	33	28	25	21									
2·6	51	49	47	44	40	36	32	28	24	21										
2·8	47	45	43	40	36	32	29	24	20											
3·0	44	42	39	37	32	28	25	20	16											
3·2	41	39	36	33	29	25	21	16												
3·4	38	35	32	29	25	22	18	13												
3·6	35	33	29	26	22	19	14													
3·8	33	30	27	23	19	16														
4·0	31	28	24	21	17	13														
4·2	29	25	22																	

were marked off, dividing the table into rectangles, each of which contained from two to six entries. The mean of the errors in each rectangle is given in table 4 below.

This table shows no serious preponderance of large errors in any area, and therefore confirms the fact that table 3 represents the observations throughout the whole region which it covers. In this connexion, it may be noted that the mean deviation between humidities in our experiments and those read from the table is $0{\cdot}3_5$ when sign is taken into account, and is 2·5 when the absolute magnitude alone is considered. It is of interest to compare the results given by this table with those from tables published previously. Only three of these are known to us, the Assmann tables⁽¹⁾, Marvin's Psychrometric Tables⁽³⁾ and those of the Meteorological Office⁽⁴⁾.

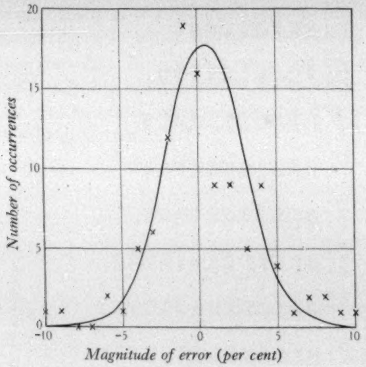


Figure 4. Distribution of deviations.

Table 4. Mean deviations of observed humidities (per cent) from the tabular values at various wet- and dry-bulb temperatures

Wet-bulb temperature (° C. below 0° C.)	Dry-bulb temperature (° C. below 0° C.)								
	2–3	3–5	5–7	7–9	9–11	11–13	13–15	15–17	17–19
5 to 6	1·7	—	0·7	—	—	—	—	—	—
6 to 7	—	0·0	1·1	—	—	—	—	—	—
7 to 8	—	−0·5	−1·0	1·0	—	—	—	—	—
8 to 9	—	—	−0·7	0·3	—	—	—	—	—
9 to 10	—	—	−0·6	2·0	—	—	—	—	—
10 to 12	—	—	—	−2·0	−1·0	−6·0	—	—	—
12 to 14	—	—	—	—	−1·0	−1·1	2·0	—	—
14 to 16	—	—	—	—	—	−1·7	−0·6	—	—
16 to 18	—	—	—	—	—	—	—	1·1*	—
18 to 20	—	—	—	—	—	—	—	—	4·2

* Including one observation at a dry-bulb temperature of 17·03° C.

In Marvin's, the basis for the calculation of the tables is stated to be the usual psychrometric formula quoted above, with B_0 taken as a linear function of the wet-bulb temperature, a condition which cannot be quite true when the wet bulb is ice-covered. The tables use a different definition of humidity from the one adopted here, since they take saturation to refer to the condition of equilibrium with ice instead of with water. In the comparison below, the relative humidities extracted from Marvin's tables have been converted into relative humidities relative to saturation with liquid water.

Table 5. Relative humidity (per cent) from present psychrometric table, and from those of previous tables

Wet-bulb depression (° C.)		Dry-bulb temperature (° C.)					
		0	−4	−8	−12	−16	−18
0	Meteorological Office	100	96	92	89	—	—
	Present paper	—	—	—	—	—	—
	Marvin	100	97	93	91	89	86
	Assmann	100	97	89	82	75	66
1·0	Meteorological Office	81	73	63	53	38	—
	Marvin	81	74	65	54	41	32
	Assmann	81	74	65	55	41	32
	Present paper	80	74	62	49	37	33
2·0	Meteorological Office	61	50	35	15		
	Marvin	63	52	38	21		
	Assmann	63	52	38	21		
	Present paper	62	53	36	24		
3·0	Meteorological Office	43	27	—			
	Marvin	45	31	12			
	Assmann	46	31	—			
	Present paper	44	32	16			
4·0	Meteorological Office	25	—	—			
	Marvin	28	20	10			
	Assmann	28	—	—			
	Present paper	31	24	17			

The only region in which any serious difference occurs is at low dry-bulb temperatures combined with small wet-bulb depressions. An examination of our experimental results in this region shows definitely a better fit for our table than for that of Assmann, as shown in table 6.

The distinctly smaller sum of the errors without regard to sign shows that table 3 represents our own observations much more closely than does the Assmann table; moreover, the sum of the errors when sign is taken into account confirms this, and shows clearly that relative to our observations, the values given in the Assmann tables are much too high in this region (about 6·4 per cent on the average). Indeed, since 21 of the errors are positive and only 3 negative, in the case of the Assmann tables, it is clear without any numerical examination that these tables give results

Table 6

Dry bulb	Wet bulb	Depression	Relative humidity			Errors	
			Assmann tables	Present tables	Observed	Assmann tables	Present tables
− 8·58	− 8·66	0·08	90	85	87	3	− 2
− 8·73	− 8·94	0·21	87	81	78	9	3
−10·70	−11·94	1·24	51	47	51	0	− 4
−10·80	−10·90	0·10	87	77	77	10	0
−10·93	−10·88	−0·05	94	81	86	8	− 5
−11·02	−11·33	0·31	80	72	66	14	6
−11·37	−11·19	−0·18	96	83	77	19	6
−11·57	−12·67	1·10	53	47	49	4	− 2
−12·16	−13·07	0·91	57	50	51	6	− 1
−12·37	−13·44	1·07	50	46	44	6	2
−13·63	−13·78	0·15	82	67	67	15	0
−13·79	−13·99	0·20	80	65	69	11	− 4
−13·99	−14·57	0·58	65	54	58	7	− 4
−14·68	−15·87	1·19	38	35	31	7	4
−15·67	−16·43	0·76	52	45	43	9	2
−16·05	−16·74	0·69	54	46	50	4	− 4
−16·06	−16·97	0·91	43	40	43	0	− 3
−16·10	−17·04	0·94	43	39	44	−1	− 5
−16·28	−17·30	1·02	40	36	27	13	9
−16·29	−17·05	0·76	51	43	42	9	1
−16·66	−17·53	0·87	44	39	45	−1	− 6
−16·75	−17·51	0·76	50	42	45	5	− 3
−17·03	−17·70	0·67	53	44	51	2	− 7
−17·36	−18·18	0·82	44	39	47	−3	− 8
−17·82	−18·22	0·40	64	51	61	3	−10
−18·22	−18·80	0·58	52	44	45	7	− 1
			Totals without regard to sign			176	102
			Totals regard being had to sign			166	−36

higher than the experimental values. In our case the negative errors are more numerous, but only in the proportion 16 to 8.

Table 7. Values of B_0P

Wet bulb depression (° C.)	Dry-bulb temperature (° C.)							
	0	−3	−6	−9	−12	−15	−18	Mean
0	—	—	—	—	—	—	—	—
0·5	0·53	0·46	0·54	0·75	0·86	0·84	0·73	0·53
1·0	0·55	0·49	0·53	0·60	0·60	0·56	0·48	0·54
1·5	0·54	0·48	0·52	0·54	0·51	0·46	0·42	0·50
2·0	0·53	0·48	0·51	0·50	0·46	—	—	0·50
2·5	0·52	0·48	0·49	0·47	—	—	—	0·49
3·0	0·52	0·47	0·47	—	—	—	—	0·49
3·5	0·50	0·46	0·45	—	—	—	—	0·47
4·0	0·47	0·44	—	—	—	—	—	0·46

If now we accept table 3 as giving the relative humidities corresponding to various wet- and dry-bulb readings, it may be used to examine the question as to how B_0 in the psychrometric formula varies with these temperatures. This is shown in table 7 below, where B_0P is tabulated, B_0 being the psychrometric "constant" and P the barometric pressure.

It will be seen that the constant tends to fall as the depression increases; as pointed out earlier, this must be so, since the constant becomes infinite at zero depression. Superposed on this variation, however, there is a considerable dependence on the dry-bulb temperature. It appears that in any horizontal line of the table, B_0P passes through a minimum and a maximum, the actual positions of which vary with the depression.

It is thus clear that the theory of the psychrometer which is found to be satisfactory when the wet bulb is covered with water, cannot be extended into the region where it is ice-covered.

§ 9. ACKNOWLEDGMENTS

This work was carried out for the Engineering Committee of the Food Investigation Board, to whom we are indebted for permission to publish. Mr M. J. Hickman rendered valuable assistance with the observations, which were of necessity carried out under conditions of discomfort, and we thank him for his interest at all stages of the work. Our thanks are also due to Mr J. K. Hardy for designing the metal enclosure and to Mr Snow for designing the special form of dew-point apparatus.

REFERENCES

(1) *Aspirations Psychrometer Tafeln* (Vieweg und Sohn, Braunschweig), 2nd ed. (1927).
(2) *Proc. phys. Soc.* **44**, 132 (1932).
(3) *Psychrometric Tables*: By C. F. Marvin, U.S. Dept. of Agriculture. Published by the Government Printing Office, Washington, U.S.A.
(4) *Hygrometric Tables*, M.O. 265, 3rd ed. (1931). Published by H.M. Stationery Office.

DISCUSSION

Prof. W. Wilson remarked that "relative humidity" is best regarded as the quantity of water per unit volume when the unit is the saturation or equilibrium quantity. Below 0° C. there are two equilibrium quantities, namely those corresponding to the solid and liquid states, and to avoid ambiguity one or the other might be adopted. He suggested that "equilibrium pressure" is a better term than "saturation pressure."

Prof. Allan Ferguson asked whether the method previously described by the authors, in which the refractive index of glycerine is taken as a measure of humidity, would be applicable to the range of temperatures dealt with in the paper?

Dr Bruce Chalmers asked whether the wet bulb is always iced below 0° C. or whether it is wetted with supercooled water. If both bulbs were dry it would be better to speak of an "evaporation hygrometer".

Authors' reply. The authors agree with Prof. Wilson that the two standards can in fact exist below 0° C. Nevertheless, to avoid unnecessary duplication of tables it seems preferable for one form of standard to be adopted universally if possible.

In reply to Prof. Ferguson: the hygrometer to which he refers could doubtless be employed at temperatures which were not low enough to freeze the glycerine solution, although it might prove much less convenient in practice than the wet-and-dry-bulb instrument.

As regards Dr Chalmers's questions: care was always taken to see that the water had frozen. Experiments in which the bulb was coated with supercooled water showed that the formula applicable above 0° C. was still valid.

It would be difficult at this stage to introduce a name such as "evaporation hygrometer" or "sublimation hygrometer" since the name "wet-and-dry-bulb" had been in use now for over a hundred years.

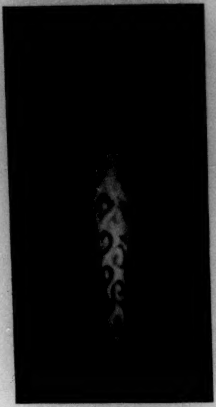
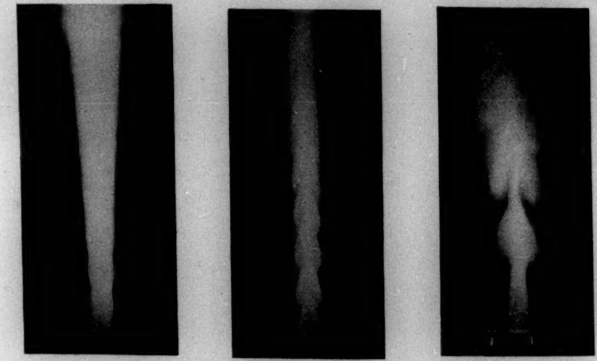


Figure 1.

Figure 2.



A 90 *cm./sec.* *B* 111 *cm./sec.* *C* 185 *cm./sec.*

Figure 4.

$n = 130$ $U = 192$ *cm./sec.* $n = 133\frac{1}{2}$

Figure 10.

534·413

ON VORTEX MOTION IN GASEOUS JETS AND THE ORIGIN OF THEIR SENSITIVITY TO SOUND

By G. BURNISTON BROWN, M.Sc., Ph.D., Lecturer in Physics,
University College, London

Received January 18, 1935. *Read May* 3, 1935

ABSTRACT. The vortices produced by sound in jets of air are examined by stroboscopic cinematography, and their velocity, growth, and angular velocity are determined. It is found possible to explain the salient characteristics exhibited by sensitive jets and the mechanism of their sensitivity. The physical processes underlying the instability of gaseous jets in general are discussed also.

PART I. THE VORTEX MOTION IN GASEOUS JETS

§ 1. INTRODUCTION

In a former paper on sensitive flames(1) it was suggested that the sinuosities observed in a stream of air when sound waves impinged upon it were due to incipient vortices travelling alternately up the opposite sides. A few observations with a stream of air issuing from a slit sufficed to show that this was the case(2). The object of the present paper is to examine the character of this vortex motion and to see what light it throws on the peculiar properties exhibited by sensitive jets. It is important, however, to note that although it is necessary, in order to be able to observe and photograph the vortices clearly, to use a column of air issuing from a slit whose width is small compared with its length, the resulting vortex-formation is not identical with that found in a jet of circular cross-section. In the first case the vortices are cylindrical and have their axes parallel to the longest length of the slit: as they grow in size they occupy the whole of the triangular space into which streams of this kind spread, figure 1, plate 1. In the case of the jet of circular cross-section the axes of the vortices are more or less semicircular and these vortices may interact, in favourable circumstances, in such a way as to cause the stream to bifurcate, one vortex being thrown to the right and the next to the left and so on, figure 2, plate 1. The experiments described in this paper were all made with jets in the form of a slit.

§ 2. EXPERIMENTAL METHOD

A large part of the apparatus used in the former investigation and described in reference (1) remained unaltered in the experiments which follow. The heterodyne oscillator and amplifier built to give a constant energy output of about half a watt, and the Rice-Kellogg moving-coil speaker were the same. So also were the pump, the large tank into which the air was discharged, and the arrangement for producing cigarette smoke. Before the air from the large tank was led through the cigarette it passed through a gas-meter, so that by means of its readings and those of a stop-watch the volume of air passing per second could be calculated. The gas-meter had a thermometer attached to it, and was connected to a pressure-gauge consisting of a mercury U-tube. The dial read up to 50 litres and each litre division was divided into ten parts: the pointer moved uniformly over the scale for the velocities required.

After passing through the cigarette the air entered an empty washing-bottle and then a long glass tube containing calcium chloride. The air was not forced to pass between the pieces of calcium chloride as this caused a variable resistance to its flow, but was allowed to flow over the top for a distance of about 50 cm. This reduced its moisture content and prevented a deposit on the jet of a sticky nicotine slime. The drying tube was connected to a wooden box 23 × 43 × 18 cm. which stood on the table facing the loud-speaker and 60 cm. distant from it. In the top of the box a brass plate was fixed containing a variable slit of length 2·36 cm. and depth 1 cm.: the width could be altered from zero up to 4 mm. and the value read off on a scale on the adjusting screw. A cardboard funnel was waxed on to the lower surface of the brass plate and hung downwards to a depth of 5 cm. inside the box. This helped to produce a more uniform flow. All joins and screw holes in the box were filled with soft wax and then coated with shellac to prevent leakage. The storage space for the cigarette smoke afforded by the box, drying tube, and washing bottle is a most important feature of the apparatus because it allows of the cigarette (6 inches long) being burnt nearly completely before any smoke appears at the jet. The screw clip controlling the flow of air through the cigarette is then closed and another clip controlling the by-pass tube is opened until the velocity is that desired. In this way the burning cigarette with its variable resistance to air flow is removed from the circuit before observation and measurement commence.

From a knowledge of the area of cross-section of the jet and the volume of air passing per second, the mean value of the velocity can be calculated. Allowance is made for the difference in pressure between the gas-meter and the box. The pressure-drop between the box and the outside air is always very small for the velocities observed and never exceeds 2 mm. of mercury. In order to keep the pressure constant in the large tank a tap attached to it was turned so as to allow a small leakage of air: by adjusting the speed of the pump, a balance between its supply and the leakage could be obtained such as to produce a pressure of a few centimetres of mercury in the gas-meter which remained constant over periods of several minutes during which photographs were taken. Care was, of course, taken to see that the escaping air did not produce draughts which could affect the jet nor sound of a frequency to which the jet was sensitive. The temperature of the air issuing from the slit never exceeded that of the air in the gas-meter by more than 2° C. The volume correction was therefore less than 1 per cent and could be neglected.

Owing to the fact that the jets examined are sensitive only to notes of low frequency and consequently of large wave-length, it was possible to surround the jet with cardboard screens, to prevent disturbance from draughts, without experiencing any trouble from reflection or diffraction. This was a very great advantage for jets of air of velocity from 100 to 200 cm./sec. are extremely susceptible to disturbance. A thick curtain of felt formed the background at a distance of 6 ft. from the speaker.

The optical system is shown diagrammatically in figure 3. Light beams from two Leitz arc lamps *A* and *B* pass through cylindrical lenses *C* and *D*, through slits in the stroboscopic disc *E* and through two more cylindrical lenses, and combine to illuminate the slit *S*. After passing this slit the two beams separate again, leaving a dark space in which the cinematograph camera *G* is placed. In this way direct light is excluded and only light which has been scattered at a small angle by the smoke enters the lens. The dull black screen *F* provides a dark background and prevents the entry of stray light.

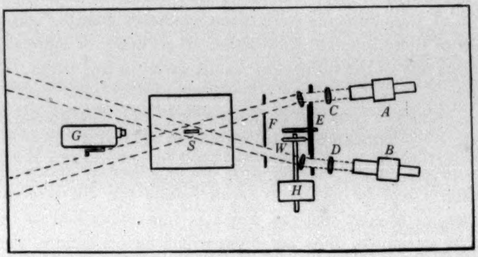


Figure 3.

The stroboscope consists of an aluminium disc of diameter 12 in. in which twelve radial slits 2 in. long are cut. The ratio of the width of the slits to the space between them, always an important matter where sharpness of definition is required, is 1 : 8. The disc is supported by an axle on ball bearings and driven by friction by a fibre wheel *W* fixed to the axle of a constant speed motor *H*. The fibre wheel bears on sheet rubber stuck on to the back of the disc, and is constructed in such a manner that its distance from the axle of the disc can be varied. Thus the rate of revolution of the disc may be altered and stroboscopic frequencies of 84–250 per sec. attained. The design was suggested by Prof. E. N. da C. Andrade; it has been used by J. W. Lewis(3) and works very satisfactorily. The position of the fibre wheel can be read off on a vernier and these readings can be calibrated in terms of the stroboscopic frequency by the use of a revolution-counter and stop-watch. The constant-speed motor *H* is controlled by an electrically maintained tuning fork of frequency 50 c./sec., and this is kept as far as possible from the jet and covered over to prevent the sound from affecting it.

The cinematograph camera was fitted with a Dallmeyer $f/2{\cdot}9$ Pentac lens and the holder allowed of the lens being pulled out sufficiently far for an object at a distance of 1 ft. to be in focus on the film. The fastest panchromatic (Eastman super-sensitive) film was used. The procedure in obtaining a series of photographs was as follows: the stroboscope was run at such a speed as to cause the vortex motion in the jet to appear nearly stationary; the stop-watch was started as the pointer of the gas-meter passed one of the litre marks on the scale and then the handle of the camera was turned slowly so that from thirty to fifty photographs were taken; finally the stop-watch was stopped at the next convenient coincidence of the pointer with a litre mark. In this way a large number of photographs of the periodic motion in its various phases at a known velocity of the stream were obtained. All the photography was done at night so that external disturbance was reduced to a minimum.

The examination of the films was carried out by means of a Leitz projector which was examined for distortion, and this was found to be negligible. The estimation of the distance between the centres of successive vortices is a matter of some skill and cannot be done with any great measure of certainty: consequently a large number of measurements were made whenever possible and the mean was taken. In view of these inherent difficulties, an accuracy of more than 5 per cent is not claimed for any of the numerical results given in this paper.

§ 3. DESCRIPTION OF CHARACTERISTIC PHENOMENA OBSERVED IN SENSITIVE JETS

Before describing the phenomena observed in jets of air rendered visible with cigarette smoke, it will be useful to recall briefly the characteristic behaviour exhibited by jets of ignited coal-gas which is fully described in the paper first referred to. With streams of illuminated gas which have a definite upper boundary the easiest method of obtaining a measure of the effect of sound upon them is to measure the height of this boundary above the jet as the frequency is increased from zero. If a graph is plotted of flame-height against frequency it is found to consist of a large number of maxima and minima within a range of frequency whose magnitude varies according to the jet used. Provided, however, that the frequency ranges of different jets have some portion in common, it is always found that the maxima and minima occur at identical frequencies for every jet. The most marked ducking (i.e. minimum of height) occurs for notes of pitch 5850, 4600, 3300, and 2400. The frequencies at which the jets were least disturbed (i.e. attained a maximum height) were 12,250, 10,850, 5300, 3500–3600, 2800 and 2200 c./sec. *These frequencies are independent of the shape of the jet,* the size of the jet, the velocity of the stream, the amplitude of the sound, and the nature of the reservoir or tubing supplying the jet with gas.* Such frequencies as could be observed with hydrogen flames agreed with those given above.

When we pass from streams of ignited gas to streams of unignited gas and streams of air in air, the range of frequencies to which they are sensitive becomes very much lower in pitch. This is to be expected from considerations of dynamical similarity together with the fact that the higher the velocity of the jet the higher the frequencies to which it is sensitive. For if we neglect all forces except those of inertia and friction, we have, by Osborne Reynolds's law of similarity, that

---

* A slight difference is found in the case of slits in which the length is large compared with the breadth. Cf. *infra*.

$$\frac{Va\rho}{\eta}=R \qquad \text{......(1)},$$

where V is the velocity, a a linear dimension, in this case the radius, ρ the density, and η the viscosity. R is Reynolds's number which is dimensionless. Thus if the fluid is changed, this formula shows that if the kinematic viscosity η/ρ is increased, the velocity, and therefore the frequency range of sensitivity, is increased also. Therefore unlit coal gas responds to slightly higher frequencies than air, and ignited coal gas to very much higher notes still, since viscosity increases and density decreases with temperature.

V, a, ρ
η, R

From the observational point of view the chief difference of behaviour exhibited by a smoke stream is that the stream has no sharp upper boundary: all that can be seen when the frequency is slowly increased from zero is that the column widens out in a fan-like manner from a certain point above the jet, and that in correspondence with the maxima and minima of height of ignited gas jets the fan closes and opens. When it closes the apex of the fan moves further from the orifice: when it opens the apex moves nearer to the orifice. The angle of the fan varies in different circumstances between a minimum of a few degrees and a maximum of nearly 90°, but if the velocity of the stream and the amplitude of the sound are kept constant the variation in angle of the fan with frequency is never more than about 20°. Consequently the variation with frequency is far less spectacular than in the case of ignited coal gas, where the height may change suddenly by as much as 20 cm. It is clear that it will be most convenient to describe the effect of sounds of different frequencies in terms of the amount of widening of the stream produced: consequently when a given frequency produces a greater widening than frequencies just above or below it, this state of the stream will be called a *maximum*, and similarly a frequency producing less widening than those just above or below, will be said to produce a *minimum*. These terms therefore stand for increased or diminished disturbance of the stream, which correspond, as will be shown, to increased or diminished vortex-growth. It is important to notice that these terms have exactly opposite meanings to those which they had in the former papers mentioned, where a maximum stood for an increased height of the flame (which corresponds to minimum disturbance) and a minimum represented a diminished height of flame (corresponding to a maximum disturbance).

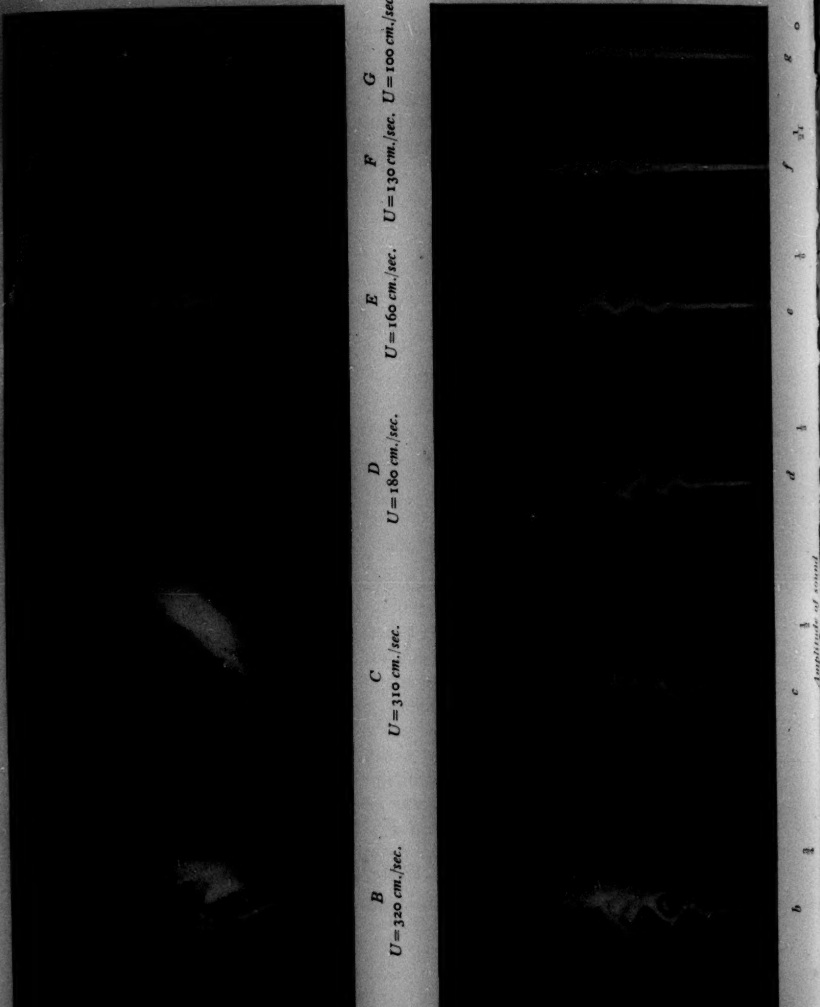
It is found that throughout the range of frequencies within which the air streams examined were sensitive there is always some vortex development, and as the frequency alters this is constantly changing; that is to say, there are in reality a large number of maxima and minima. It is of interest to see whether the frequencies for a circular jet are exactly the same as those for a slit, and to determine this a slit and a glass tube were set up close to one another, so that their behaviour could readily be observed as the frequency was changed. It was found that their maxima and minima coincided in almost every case, and only one frequency was found which represented a maximum in one case and a minimum in the other. The relative effect, however, differed considerably, and consequently in table 1, which gives only well-marked maxima and minima, the frequencies in one column do not always coincide with those in the other.

Table 1. Frequencies (c./sec.) for jets of air in air

Slit		Circular orifice	
Maxima	Minima	Maxima	Minima
76	—	84	—
94·3	—	—	—
97·6	—	—	—
126	—	126	112
130	133·5	—	—
138	136	145	—
163	—	—	—
175	—	175	—
190	—	190	—
210	—	215	290
346	384	310	390
373	390	—	—
455	—	455	420
510	—	—	475
660	—	660	510

The general characteristics of the vortex motion can be seen by examining figure 1, plate 1, which corresponds to a slit-width of 1 mm. The stream remains straight-sided in the shape of a wedge of very small angle up to a certain height above the orifice, and then a bulbous disturbance appears first on one side and then on the other. This travels upwards and very soon shows that it is an incipient vortex by forming a hook-like portion which spreads downwards towards the orifice and then curls round inwards and continues in a nearly circular path, the gradual expansion of the outer portions making this circular motion possible. As the vortices travel up the stream their shape becomes markedly triangular and then finally quadrangular. These quadrangular vortices have unequal sides and angles and pack together in such a way as to produce the familiar wedge-shaped spreading. It will be noticed, further, that above a certain height (marked A and A' in figure 1) the vortices no longer entrain outside air, the boundary being a closed surface from this height upwards. It is at this point also that the rate of rotation in the vortex slows down and becomes, as far as one can observe, ultimately zero. The entrainment of outside air can be shown very strikingly by having the stream invisible and allowing cigarette smoke to approach it at various heights. If the cigarette smoke is admitted into the stream anywhere above A and A' it streams away upwards and there is no sign of circulation. To get smoke into the region marked B the smoke must be admitted below AA', and to fill it completely must be entrained close to the orifice.

The above description holds in the case of streams varying in width from


A
U = 380 cm./sec.
B
U = 320 cm./sec.
C
U = 310 cm./sec.
D
U = 180 cm./sec.
E
U = 160 cm./sec.
F
U = 130 cm./sec.
G
U = 100 cm./sec.
a
1
b
¾
c
½
d
¼
e
⅛
f
1/24
g
0
Amplitude of sound

0·14 mm. up to 4 mm., which was the range investigated, but at both limits the motion becomes difficult to observe. For narrow jets the velocity has to be high and the vortices consist of very tenuous filaments, while in wide jets the velocity is low and the vortices are very thick and diffuse; compare figure 4*C*, plate 1. A width of about 1 mm. represents an optimum value, and this was used whenever possible throughout this research. The distance between the centres of successive undulations or vortices *on the same side* of the stream will in what follows be referred to as the wave-length λ of the vortex motion: it is the distance moved by a vortex in a time corresponding to the period of the sound waves affecting the jet. This distance varies slightly from vortex to vortex and tends to increase with increasing height above the orifice. To estimate it is easy when the vortex development is small, but becomes more and more difficult when the development is great, as can readily be seen by examining figure 5 (*A*–*C*), plate 2. In every case as many measurements were made as the number and distinctness of the film photographs allowed.

λ

The variation of wave-length with velocity of the stream and with frequency of the sound and the question of the angular velocity of the vortices and their rate of growth and other characteristics will be dealt with under separate headings. It will be convenient however to consider under the two first headings some results which do not depend on stroboscopic cinematography.

§4. VELOCITY FOR TURBULENCE

It is well known that streams of fluid are only markedly sensitive to the vibrations of sound when at, or very near, the point of breaking down into turbulent flow, although after this point is passed, it is still possible to 'force' a periodic vortex formation on the otherwise completely irregular turbulence(1), and periodic disturbance can also be produced at very much lower velocities, as will be seen later. A graph in which the velocity for turbulence is plotted against the slit-width is given in figure 6. The stream was taken to be on the point of turbulence when a flickering appearance first became visible as the velocity was increased. The flickering appearance is due to irregular vortices travelling rapidly up the column. The Reynolds lower critical velocity for turbulence in tubes of diameters equal to the widths of the slit is also shown.

The graph shows very clearly that the phenomenon of acoustical sensitivity in jets is not connected with turbulence already existing in the stream when it issues from the orifice, since the velocity for sensitivity is far below that for turbulence in the orifice itself. Its non-hyperbolic form is also noticeable. In what follows the *critical velocity* will always refer to that for maximum sensitivity to sound: the critical velocity for turbulence inside the orifice itself, which is so much higher, will be called the ***Reynolds critical velocity***.

§ 5. LIMITING VELOCITIES FOR VORTEX DEVELOPMENT AT DIFFERENT FREQUENCIES

When the velocity of the stream issuing from a jet of given width is increased until it becomes sensitive, it is always found that this sensitivity only appears for frequencies above a certain minimum value. The curves of minimum frequency against velocity for jets of various widths are given in figure 7. It will be observed that they are very nearly linear. The points on the graph were obtained by setting the condenser of the variable oscillator to 10°, 20°, 30°, etc. and measuring the velocity of the stream for the first appearance of vortex *development*, i.e. for the first visible trace of increased angular spreading of the column. This is an important point, because if we are satisfied with vortex-formation which may disappear within a few millimetres from the orifice, then very much lower velocities will suffice. For instance, with the aid of the stroboscope, a 1-mm. jet will show vortices symmetrically placed with respect to the median plane at a velocity of 42 cm./sec., but these disappear in a distance of 3 or 4 mm. The velocities given in figure 7, therefore, are those sufficient to develop the vortices once they are formed at the orifice, a condition necessary for the stream to exhibit sensitivity which is readily visible to the unaided eye.

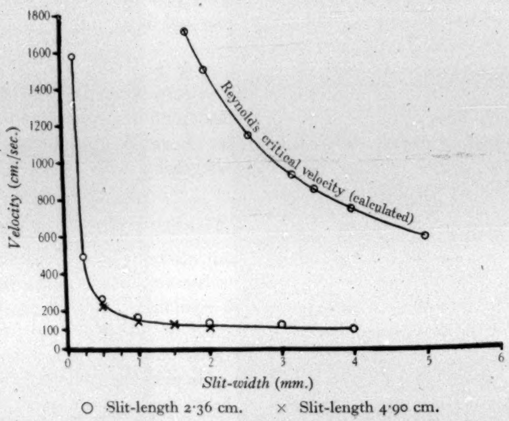


Figure 6. Mean velocity for turbulence.

It happened that one of the readings to which the condenser was set (126 c./sec.) was a very marked maximum and in these cases the minimum velocity is always lowered, so that a marked drop in the curves appears at this frequency. Clearly, if the condenser-settings had included other marked maxima and minima the curves would not be smooth as shown, and these must therefore be regarded as representing average values. It will be noticed that the points for jets of width 1 to 4 mm. lie very close together, but from 1 mm. downwards the difference becomes more and more marked. The effect of the frequency 126 c./sec. on the smallest jet of all, 0·14 mm., is also very noticeable.

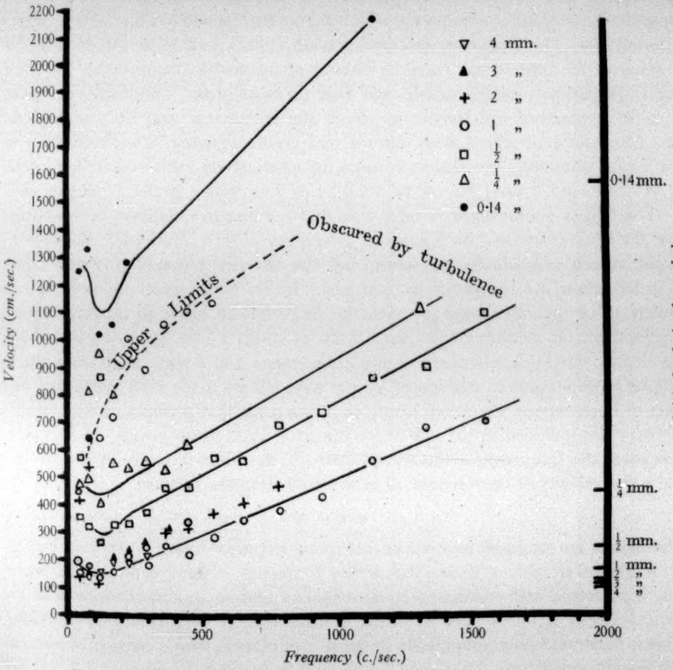


Figure 7. Velocity range within which a given frequency produces visible vortex development.

Attempts to measure the upper limit of the velocity for development of a given frequency were largely frustrated by turbulence setting in, but the few observations which could be made are recorded in the graph.

The maximum and minimum distance apart of the vortices cannot, of course, be calculated from figure 7 without a knowledge of the vortex velocity relative to the mean velocity of the stream at the various frequencies. This is a matter which will be referred to later.

§6. VARIATION WITH RESPECT TO THE STREAM-VELOCITY OF THE RELATIVE VORTEX-VELOCITY

A number of films were taken of the vortex motion in streams of varying width and velocity and for different frequencies. A frequency was chosen to represent the general case; that is, a frequency which did not correspond to a marked maximum or minimum. The frequency selected was 97 c./sec., and with the stroboscope running at this rate it was possible without alteration to examine also the effect due to frequencies double, treble, and four times as great. The results are given in table 2, together with others in which the frequency was not a multiple of 97 c./sec., some of which were minima and others maxima. The consistency of the figures obtained show that in considering phenomena exhibited at frequencies of 97 c./sec. and octaves thereof we shall be dealing with a perfectly general case.

The figures are arranged in table 2 so that the first five columns in most cases give the results obtained for a constant frequency of 97 c./sec., while the velocity of the stream is gradually increased, and the last five columns give the values recorded when the frequency is increased. In the latter case the velocity had usually to be increased together with the frequency in order to obtain adequate development for measurement. Since a slit of width 1 mm. gives, on the whole, the clearest vortex-development of any in the range 4 to $\frac{1}{4}$ mm., such extra observations as were made to test special points were always made with this width, and a set of observations which refer only to frequencies that produce maximum disturbance are included in the last five columns of the 1-mm. group. In this table n denotes the frequency, λ the wave-length, U the mean velocity of the stream, and u the velocity of the vortices. u is obtained from the relation

n, λ, U

u

$$n\lambda = u \qquad \text{......(2).}$$

The figures are arranged in order of increasing stream-velocity.

A glance at table 2 shows that if the frequency is kept constant the wave-length increases with increasing stream-velocity and so in consequence does the vortex-velocity u, but in such a manner that the relative vortex-velocity decreases. This is illustrated photographically in figure 5 of plate 2, which consists of natural-size enlargements from a film, and by a graph in figure 8. Figure 5 (*A–G*) shows the successive stages exhibited by the vortex-formation in a stream 1 mm. wide, for sound of frequency 126 c./sec. and constant amplitude, with the mean velocity varying from 380 down to 100 cm./sec., which, as can be seen from figure 6, is well below the turbulent velocity of approximately 150 cm./sec. Commencing with the lowest velocity, it can be seen that near the orifice the vortices are level with one another, i.e. are symmetrically disposed with respect to the median plane of the stream. This is well shown in the case of wider jets, as in figure 4 (*A–C*), plate 1. As the vortices proceed, one of them is pushed ahead until they become disposed alternately, a disposition which they retain until they are finally dispersed. This stage can be clearly seen; the vortex on the right-hand side is being pushed ahead, and in consequence its velocity momentarily is rising from 45 to 63 cm./sec.

Table 2

Width of slit	Reference number of film	n (c./sec.)	λ (cm.)	U (cm./sec.)	u/U	Reference number of film	n (c./sec.)	λ (cm.)	U (cm./sec.)	u/U
4 mm.	15A	97	0·80	85·6	0·91	16A	97	0·88	144	0·59
	15B	97	0·86	96·2	0·87	16D	95	0·93	150	0·59
	15C	97	0·90	132	0·66	16E	95	1·04	178	0·56
	15E	97	1·20	221	0·53	16B	194	0·67	189	0·69
	—	—	—	—	—	16C	291	0·64	212	0·88
3 mm.	12A	97	0·62	74	0·81	—	—	—	—	—
	12B	97	0·77	89	0·84	13A	97	0·81	111	0·72
	12C	97	0·84	130	0·62	13B	194	0·71	225	0·61
	12D	97	0·97	179	0·53	13C	291	0·60	249	0·71
	12F	97	1·10	221	0·48	13D	388	0·50	314	0·62
2 mm.	6A	97	0·66	97	0·66	—	—	—	—	—
	6B	97	0·71	110	0·63	7B	97	0·72	110	0·64
	6E	97	0·71	146	0·47	7A	97	0·71	149	0·46
	6C	97	0·98	205	0·46	—	—	—	—	—
	6D	97	1·03	225	0·44	7E	194	0·56	156	0·70
	6F	194	0·61	193	0·61	7C	194	0·59	195	0·59
	—	—	—	—	—	7D	194	0·58	206	0·54
	—	—	—	—	—	7F	291	0·51	242	0·60
	6G	388	0·55	388	0·55	7G	388	0·51	350	0·56
1 mm.	8A	97	0·52	107	0·47	17B	94·3	0·59	164	0·34
	8B	97	0·58	185	0·30	17D	138	0·50	191	0·36
	8C	97	0·70	237	0·29	21D	130	0·55	192	0·37
	22K 2	126	0·32	90	0·45	21D	133·5	0·55	192	0·37
	22G 2	126	0·36	100	0·45	17G	190	0·45	249	0·34
	22F 2	126	0·54	125	0·54	—	—	—	—	—
	22E 2	126	0·57	160	0·45	17F	175	0·47	258	0·32
	22F 3	126	0·63	200	0·40	—	—	—	—	—
	17C	126	0·53	250	0·27	17H	210	0·45	276	0·34
	8D	194	0·51	253	0·39	17E	163·5	0·51	283	0·30
	8E	291	0·41	282	0·42	9B	373	0·44	444	0·37
	8F	388	0·46	374	0·48	9C	384	0·44	430	0·39
½ mm.	10A	97	0·51	224	0·22	11A	97	0·63	309	0·20
	10B	97	0·61	276	0·21	11B	194	0·45	392	0·22
	10D	97	0·64	375	0·17	11C	291	0·40	451	0·25
	10E	97	0·69	396	0·17	11D	388	0·43	620	0·27
	—	—	—	—	—	11E	485	0·37	700	0·26
¼ mm.	14A	97	0·41	581	0·068	14D	97	0·56	665	0·081
	14B	97	0·62	894	0·067	14E	194	0·53	922	0·11
	14C	97	0·88	1640	0·052	—	—	—	—	—

The increased flattening of the vortex produced by this excess velocity relative to the air outside is also very marked, and results, as we should expect, in increased circulation, so that when this vortex finally settles down in its place it shows greater development than the originally stronger ones on the left-hand side. This interesting phenomenon holds throughout the whole series of films, and the reason why the vortices on the left-hand side are originally the stronger and able to push the others ahead is no doubt that the velocity of the stream on this side was very slightly greater than on the other. This was discovered in some unpublished experiments on edge tones, in which the edge of a wedge which bisected the stream at the orifice no longer did so exactly when raised to a height of several centimetres above it. The photographs *F*, *E* and *D* show the development produced by increasing the velocity to 110, 160 and 180 cm./sec. respectively. A still further increase in velocity produces the appearance shown in *C*, *B* and *A*, where the sound is able to force a periodic motion on the general turbulence.

If now we examine the graph in figure 8, where the values of u/U for jets of widths from 4 to $\frac{1}{4}$ mm. are plotted against the stream-velocity, we see that u/U remains constant or may even rise slightly for low velocities* and then falls, rapidly at first in the case of wide jets and more slowly in the case of narrow ones, until for higher velocities it tends to become constant again. A comparison of these curves with the respective films shows that the initial flat portion corresponds to the stage in which the vortices are changing from the symmetrical to the alternate position, and the velocity is below the critical value. The portion of the curve in which u/U drops corresponds to the stage where the vortices are alternate and the development is in the stage *G* to *D* of figure 5. Finally the later parts of the curves, where the value of u/U seems to tend to constancy once more, correspond to what we may call the excessive-development stage illustrated by *A*, *B* and *C* in figure 5, in which the velocity is about double that of the critical velocity for turbulence without sound.

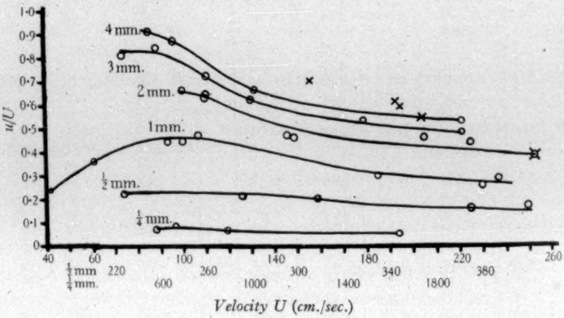


Figure 8.

It is clear that vortices may travel in streams of air with relative velocities which vary within large limits according to the width of the jet, and there is no tendency

---

* Figure 8 and table 2 refer to *developed* vortex-formation, which means that the column of air is visibly disturbed when viewed in the normal way. With the aid of stroboscopic illumination, however, it is possible to discover, at much lower velocities, vortex-development in the symmetrical stage which dies away within a few millimetres of the jet. This is very difficult to observe, and was clearly seen only in the case of the 1-mm. slit: an estimate of the wave-length λ was made and the resulting values of u/U were plotted in figure 8, giving the first two points on the 1-mm. curve which clearly rises to a maximum just before the alternate stage sets in.

K, b

to reach a limiting value of half that of the stream as calculated by Rayleigh(4). In this calculation, applicable to "sinuous disturbance", Rayleigh gets u/U to be $1/(1+\cosh Kb)$, where $K=2\pi/\lambda$ and b is half the width of the slit. This is a curve rising from the origin and approaching asymptotically the value $\frac{1}{2}$. If suitable values of u/U and Kb are taken from table 2 and plotted on the same graph, the curves have only one point in common, where they cut one another, and the limiting value of u/U is 0·7. That these results are not in agreement is hardly surprising in view of the non-uniform velocity-distribution in the jet and the large entrainment of air by the vortices in the alternate position, a condition not allowed for by Rayleigh, who neglected the effects of viscosity. It should be noted, however, in this connexion that E. G. Richardson, using a hot-wire method, found good agreement with Rayleigh's formula in the case of the vortex formation in edge tones(5).

Even if the agreement is not satisfactory in the case of vortices in the alternate position, it might seem reasonable to suppose that Rayleigh's calculation for the case where the disturbance is symmetrical about the axes of the jet would apply to the first stage in which the vortices are opposite one another and there is little or no entrainment of air (cf. figure 4 *A*, plate 1). But in this case Rayleigh finds u/U to be $1/(1+\tanh Kb)$; and this is a curve which starts at the value unity and falls off, approaching asymptotically the value $\frac{1}{2}$. This is, of course, completely contrary to the results obtained, which indicate that the curve starts at or near the origin and approaches asymptotically the value unity.

d

Another comparison of theoretical results is of interest in the case of the Kármán vortex-street. In table 3 are given values of d/λ in cases where d, the distance between the rows of vortices, remains constant and is equal to the width of the stream. It is clear that there is no evidence of d/λ tending towards Kármán's value 0·28.

Table 3

Slit-width d (mm.)	Film	λ (mm.)	d/λ
¼	14D	5·6	0·05
½	10B	6·1	0·08
1	8B	5·8	0·17
2	6B	7·1	0·27
3	12D	9·7	0·31
4	15E	12·0	0·33

The crosses marked in figure 8 correspond to values of u/U for a frequency of 194. The cross on the extreme right is for a 1-mm. jet and lies on the curve for 2-mm. jets. The four remaining crosses are for the 2-mm. jet and one is seen to lie on the 4-mm. curve; the remaining three however indicate that we cannot say that the vortex-production in a given jet for a given frequency is dynamically similar to that in a jet of half the size and double the frequency, as we might be led to suppose if the hyperbolic relation between the size and the velocity assumed in equation (1) were true and might deduce from Rayleigh's formula above.

§ 7. VARIATION OF RELATIVE VORTEX-VELOCITY WITH FREQUENCY

We may now consider how the relative vortex-velocity u/U varies with the frequency of the sound when the velocity is kept constant. If we select from table 2 values for which the velocity is the same or nearly the same, we find that the relation between u/U and n is very nearly linear and since U is constant we may write

A, *B*

$$u = An + B \text{ where } A \text{ and } B \text{ are constants} \qquad \ldots\ldots(3),$$

$$(U = \text{const.})$$

And since $n\lambda = u$ we have the further relation

$$\lambda = \frac{B}{n} + A \qquad \ldots\ldots(4).$$

Figure 9 illustrates these results in the case of velocities in the neighbourhood of 230 cm./sec. for jets from 1 to 4 mm. in width.

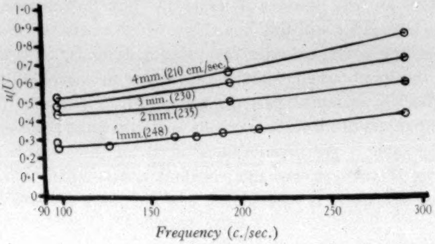


Figure 9.

§ 8. VORTEX-GROWTH

If the critical velocity for turbulence is not exceeded, the boundaries of a stream of air are nearly straight and diverge slightly as the distance from the orifice increases owing to diffusion and the consequent falling off of the velocity. If, however, vortices are produced either by excessive velocity or by sound, these vortices grow as they pass up the column and the boundaries are then no longer approximately linear but curve outwards in the manner shown in figure 5 (*C*–*G*), and if these are examined it will be observed that for velocities up to about 160 cm./sec. the vortices, although they grow, remain attached to one another: further increase in velocities causes them to break up, a process which is beginning in figure 5 (*D*) and which is very evident in the remaining photographs (*A*, *B* and *C*) where the vortices literally burst and filaments of smoke spread outwards to such an extent that in some cases they get entrained by the following vortex. This is well shown in figure 5, especially in (*A*), and occurs most markedly on the right side where, as we have seen, the vortex-development is greatest. It is interesting at this point to compare this growth with that which takes place in circular jets (figure 2, plate 1) where, as has been remarked, the vortices on separating may, in favourable circumstances, be flung into two paths inclined at an angle, the intervening space being free from smoke. In figure 2 the forking is not quite perfect and filaments extending across this space may be discerned.

The vortex-growth in wide jets follows the same course, with the important difference that the development may not reach the stage at which the vortices become alternate before the velocity is so high that the resulting general turbulence masks any further development (cf. figure 4, plate 1).

In the lower part of figure 5 (*a–g*), plate 2, is shown the vortex development as the amplitude of the sound of frequency 126 c./sec. is decreased from its full value to zero for a 1-mm. slit and the velocity is constant at 150 cm./sec. It is interesting to notice in comparing photographs in which the angular spread is the same, e.g. *D* and *d* or *E* and *e* in figure 5, that the development begins much further from the orifice in the case of the reduced-amplitude series. Consequently, although we can obtain a given angular spread by adjusting either the velocity or the amplitude of the sound, if we consider the jet as a whole its condition is unique and depends on both velocity and amplitude. Another noticeable feature is the increase in thickness of the vortex filaments with increase in amplitude of the sound. Figure 5 *g* shows the stream with the loud-speaker shut off: owing to the velocity being the critical velocity of 150 cm./sec. there are signs of slight vortex-development initiated probably by sound from the rotating stroboscopic disc.

We may now consider in more detail the conditions which control the angular spread of the stream when the amplitude of the sound is kept constant at its maximum value. A hint in this direction is given by the second group of figures for the 1-mm. slit given in table 2. All these results refer to cases in which the velocity has been adjusted so that the vortex-development is as great as possible and at the same time is clearly defined in the upper portions, and it will be seen that the value of u/U is practically constant at about 0·35. This indication that the stage of vortex-development is dependent on u/U is also borne out by the following tables in which values for slits of different widths are compared for three different stages: (A) minimum sinuous development, in which the vortices are alternate but do not increase in size appreciably, giving to the column a sinuous outline; (B) small angular development of about 20°; and (C) maximum angular development (about 60°) in which the vortices break up so that they are no longer clearly defined in the upper region. The velocity for this stage is that at which the angular spread first reaches 60°. The figures referred to in table 2 indicate that there is a linear relation between n and U for constant angular development.

In attempting to see what effect was caused by the variation in the slit-width, it was found that for the widths 1 mm. to 4 mm. $u/Ud^{\frac{1}{2}}$ was approximately constant and values of this function are given in the table. The widths $\frac{1}{4}$ and $\frac{1}{2}$ mm. do not give values in agreement, and, considering the difference between the 1- to 4-mm. group and smaller slit-widths already noticed in figure 7, this is perhaps not to be

Table 4

Slit (mm.) ...		$\frac{1}{4}$	$\frac{1}{2}$	1	2	3	4
A. Minimum sinuous development							
Frequency 97 c./sec.	Velocity	581	224	107	97	89	85
	λ (cm.)	0·41	0·51	0·52	0·66	0·77	0·80
	u/U	0·07	0·22	0·47	0·66	0·84	0·91
	$u/Ud^{\frac{1}{2}}$	0·14	0·31	0·47	0·47	0·44	0·46
	Film	14A	10A	8A	6A	12B	15A
Frequency 291 c./sec.	Velocity	—	—	282	242	249	212
	λ (cm.)	—	—	0·41	0·51	0·60	0·64
	u/U	—	—	0·42	0·60	0·71	0·88
	$u/Ud^{\frac{1}{2}}$	—	—	0·42	0·42	0·41	0·44
	Film	—	—	8E	7F	13C	16C
B. Small angular development (20°)							
Frequency 194 c./sec.	Velocity	922	392	253	206	225	189
	λ (cm.)	0·53	0·45	0·51	0·58	0·71	0·67
	u/U	0·11	0·22	0·39	0·54	0·61	0·69
	$u/Ud^{\frac{1}{2}}$	0·22	0·31	0·39	0·38	0·35	0·35
	Film	14E	11B	8D	7D	13B	16B
C. Maximum development							
Frequency 97 c./sec.	Velocity	1640	375	237	205	221	221
	λ (cm.)	0·88	0·64	0·70	0·98	1·10	1·20
	u/U	0·05	0·17	0·29	0·46	0·48	0·53
	$u/Ud^{\frac{1}{2}}$	0·10	0·24	0·29	0·32	0·28	0·27
	Film	14C	10D	8C	6C	12F	15E

expected. However, if we restrict ourselves to the 1- to 4-mm. group we may state that

$$\left.\begin{array}{l}\text{Minimum sinuous development occurs when } u/Ud^{\frac{1}{2}} \simeq 0{\cdot}43 \\ \text{Maximum angular development occurs when } u/Ud^{\frac{1}{2}} \simeq 0{\cdot}29\end{array}\right\} \quad \ldots\ldots(5).$$

It must be noted that the examples given in the above tables are the best that could be selected from the various films, and as the films were not taken with this special object in view the values are not always entirely consistent. A certain latitude also inevitably exists in deciding when two streams are in the same stage of development. Thus the table above indicates a fall of u/U with increasing frequency, whereas the figures in table 2 indicate a slight rise. We may therefore conclude that the values of $u/Ud^{\frac{1}{2}}$ are approximately independent of the frequency.

The formula (5) for maximum development gives a fair approximation to the asymptotic value to which the curves shown in figure 8 approach.

§ 9. ANGULAR VELOCITY OF THE VORTICES

It is clear that when the vortex-formation appears stationary in the stroboscopic illumination, as it is for instance in the photographs, any given vortex represents the stage which the next vortex below it on the same side would reach in a time equal to the reciprocal of the frequency. Consequently if we note the angle turned through by the leading filament of the vortex between these two positions and call this α, then its rate of rotation ω is given by

α, ω

$$\omega = \alpha n.$$

Now the most surprising fact emerges from the study of all the hundreds of feet of film taken and from visual observation made specially for the purpose, that *the angle is always about* 180° *or* π *radians*. It is difficult to estimate exactly since the leading edge of the vortex filament is not always very distinct, but in every case, for jets of all sizes, for all velocities, and for all frequencies up to 600 at least (above this frequency the vortices are very small and difficult to observe clearly) the angle α is very close to the value 180°. We have therefore the fundamental relation

$$\omega = \pi n \text{ radians per sec.} \qquad \ldots\ldots(6),$$

that is to say, *the angular rotation of the leading filament of the vortices is directly proportional to the frequency of the sound.*

It will be evident at once that this striking fact contains the secret of the most puzzling result of the former research on sensitive flames(1), viz. that whatever the size of the jet, the velocity of the stream, the amplitude of the sound, and the nature of the reservoir or tubing supplying the jet, the frequencies for maximum and minimum disturbance for a given gas are fixed and unalterable. For now that we have discovered that the angular velocity depends only on the frequency and is independent of the velocity and slit-width, and moreover since it is the *only* feature of the phenomenon of which this can be said, we are led to the conclusion that *some rates of angular rotation of the vortices are more favourable to vortex-development than others.*

And as soon as we make this statement, the question at once arises whether the statement applies to vortex-formation in the symmetrical as well as in the alternate position. When this point was tested experimentally, it was found that as long as the vortices remain symmetrically placed and do not attain the alternate position, there is no sign of any differential effect with change of frequency. We may therefore amend our former conclusion to the following: *some rates of angular rotation are more favourable to the interacting growth of vortices in the alternate position than others.*

It is evident that at high frequencies the rate of rotation is very great. In the case of the highest frequency in which the rotation could be accurately observed, viz. about 600, we have the leading filament of the vortex rotating at approximately 300 rev./sec. If formula (6) holds in every case, then we should have, for the limiting frequency available for a 1-mm. stream which is about 1800 c./sec. (mean velocity 700 cm./sec.), a rate of rotation of 900 rev./sec. Again if the same relation holds for burning coal-gas, jets of which are in some cases sensitive to frequencies of over 20,000, the rate of rotation must be as great as 10,000 rev./sec. The fact that the angular velocity is constant for the leading filament of these spiral vortices must not lead us to infer that it is also constant for the outer portions. In fact, a little consideration will convince us that this cannot be the case, and that the motion is a very complicated one. This consideration of the circulation and its origin will be given in a later section.

§ 10. THE DIFFERENCE BETWEEN MAXIMA AND MINIMA

A series of photographs was taken in two instances where the slightest turn of the condenser dial of the variable oscillator changed a maximum vortex-development into a minimum. These were for frequencies of 373 (max.) and 384 (min.) and for 130 (max.) and $133\frac{1}{2}$ (min.) for constant velocities of 430 cm./sec. and 192 cm./sec. respectively. The last pair is illustrated in figure 10, plate 1, and two horizontal lines have been superimposed to aid in comparing the two photographs, which were very carefully selected so as to exhibit the vortex-formation in the same phase in each. In both cases it is clear that no fundamental difference between the vortex-formation at a maximum and at a minimum exists. A change of $2\frac{1}{2}$ per cent in the frequency produces no observable difference in the wave-length λ and consequently none in u/U, since the mean stream-velocity U is the same in each pair. At the same time the difference in vortex-development in the two cases is very marked, as is also the subsequent angular spreading. Now, as we already know, if we alter the slit-width, or if we alter the velocity, we should alter both λ and u/U, and yet the frequency 130 c./sec. will always produce better development than $133\frac{1}{2}$, and 373 than 384. So we are forced to the conclusion put forward in the last section that the rate of rotation of the vortices in the alternate position has a very marked effect on their development, since this is the only physical phenomenon that depends directly on the frequency and is independent of λ, u/U, the slit-width, and the amplitude of the sound.

PART II. THE ORIGIN OF THE SENSITIVITY OF GASEOUS JETS TO SOUND

§ 11. DISTURBANCE OF THE STREAM BEFORE IT LEAVES THE ORIFICE

The possibility that turbulence may be already present in the stream when it emerges must be ruled out because the velocities are so much below the Reynolds critical velocity; e.g. in 1-mm. jets this velocity is about 3000 cm./sec., whereas the stream begins to flicker at 150 cm./sec. and is sensitive to loud knocks at 130 cm./sec., and further, as has been mentioned, traces of vortex-production can be seen with the stroboscope at velocities as low as 42 cm./sec. Clearly, then, in smooth-lined jets we are not dealing with vortex-motion in the stream before it leaves the orifice. Further, jets stuffed with cotton wool are still sensitive and show the same maxima and minima although the frequency range for sensitiveness is usually altered, as was shown in the former paper.* The effect of periodic disturbance in the supply tubing was tried by placing a whistle in the tubing, and this causes the critical velocity to be lowered and the sensitivity to be much impaired.†

It is of interest to note that as the velocity is progressively increased from the sensitive value the wedge-shaped spreading gradually approaches the jet, the last movement to the jet itself taking place with a sudden jump. The velocity at which this jump takes place and at which the wedge-shaped spreading commences at the jet itself was found to agree well with the Reynolds critical velocity. This is to be expected, since if vortices already occur in the stream they need no space for development, and the wedge of entraining air can begin at the jet. On the other hand, if the disturbance commences at the jet we should expect a short length of apparently undisturbed column in which the vorticity is increasing to the stage at which a filament spreads out and entrains air as shown in figure 1, plate 1.

---

* Reference (1), figure 9, p. 176.

† Reference (1), p. 166.

The case of a jet through which air is travelling at a speed above the Reynolds critical velocity and in which, in consequence, the spreading begins at the jet, has been dealt with mathematically by Tolmien(6), who adapted a method due to Prandtl(7), and the results are found to agree well with experimental values obtained at Göttingen. No mathematical treatment of the earlier stages of jet turbulence has, however, been successful.

§ 12. VIBRATION OF THE JET

In the case of the symmetrical stage in large jets it is easily observable that the column alternately swells and contracts as it leaves the jet, and the magnitude of this lateral motion is of the order of a millimetre, whereas the jet is, as far as can be seen, quite stationary; cf. figure 4 *A*, plate 1. An experiment was made with an Amplion diaphragm speaker to which a cylindrical funnel 5 cm. long was attached, so that when the amplitude was very small the only motion of the air was in the direction of the axis of cylinder. When this speaker unit was placed within a few centimetres of the jet and with its axis horizontal and at right angles to the length of the slit, alternate swelling and contraction of the stream at the orifice could be seen. If the loud-speaker was removed to 10 cm. no effect was discernible owing to the extreme feebleness of the sound, which could only just be heard by placing the ear at the mouth of the cylindrical opening. The sensitivity of jets to sounds is of the order of that of the human ear. In this position the diaphragm and the outer side of the slit were put in contact by means of a light metal rod, and thus the vibrations could be transmitted directly, but no trace of disturbance was visible. The same amplitude of vibration of the air had produced marked effects at a distance of a centimetre or two, and so we can conclude that vibration of the jet is not the cause of sensitivity in gaseous streams.

§ 13. PRESSURE EFFECTS

It is already well known that jets are most sensitive, in almost every case, at the antinodes in stationary waves. This fact must be put beside the facts established in the former paper, that the insertion of resonating chambers in the orifice or supply, and the filling of the orifice with cotton wool, do not materially alter the sensitivity. These facts all lead to the conviction that the sensitivity is not a pressure effect but is in some way due to motion of the air and must therefore be directional. This point was conclusively demonstrated by means of the Amplion unit mentioned above: it was found that the maximum vortex production occurred when the loud-speaker caused the air to vibrate horizontally and at right angles to the length of the slit, and that no effect whatever could be found when the Amplion loud-speaker faced downwards into the orifice. Further, it was possible, with wide jets in stationary waves, to find positions in which vortex production occurred *only on one side* of the stream. Clearly, then, the disturbance of the jet is a lateral one, and is a maximum at right angles to the slit length and is not due to vibrating of the air in and out of the orifice; although we do, of course, know that symmetrical swellings in a column of gas can be produced by vibratory movement of the air in a direction parallel to the stream. This, for instance, is produced in the well-known singing-flame experiment.

Periodic disturbance of a jet of gas by vibration parallel to the direction of flow is therefore possible and does occur in at least one well-known case, but apparently this is not the type of disturbance produced by sound, and probably is only effective for large-scale motion of air such as occurs in resonating pipes.

§ 14. DIRECTIONAL EFFECTS

Experiments with the Amplion unit and very low intensity (to avoid reflection) showed that vortex motion was produced in the sides of the stream when the loud-speaker faced the side of the slit, and in the ends when the loud-speaker faced the ends, but not *vice versa*. Vortices were therefore produced only with axes perpendicular to the direction of sound, i.e. perpendicular to the direction of motion of the vibrating air. The vortex motion was always greatest when the motion of the air was perpendicular to the edge of the stream. Circular cross-sectioned jets showed, as would be expected, approximately equal sensitivity in all directions, and the vortex-production and forking occurred in a vertical plane containing the direction of motion of the vibrating air. Needless to say, very few circular jets give an accurately uniform flow over their circumference, but differences in velocity, if they occur at opposite ends of a diameter, favour the production of marked forking, and such jets exhibit marked directional properties.

§ 15. SENSITIVITY ONLY AT THE JET

It is well known that the sensitive region in a gaseous stream is in the immediate neighbourhood of the orifice. This can be demonstrated by lowering a shield near the jet and between it and the loud-speaker: no change in the vortex-formation can be observed until the shield cuts off the sound waves from the last millimetre or so at the orifice. This also shows that once the vortices are started the sound has no effect on their further development. It was also possible to demonstrate, by moving the Amplion unit about, that nothing ever occurred unless the loud-speaker pointed at the column close to the jet. We may therefore conclude that a jet is essential to sensitivity, and gaseous streams by themselves are insensitive. To see more clearly why it is a fundamental requisite, we may first attempt to define what is meant by a jet.

§ 16. WHAT IS A JET?

A jet is essentially a *solid boundary*, and this means that any stream of gas passing it emerges with those portions of it which have been in contact with the solid boundary practically at rest. And it is to this outer sheath of slow-moving molecules in the neighbourhood of the jet that the sensitivity to sound is due, for these molecules start to diffuse outwards, and since diffusion is a molecular process we shall expect that such a process will be affected by other molecular motions such as the periodic to-and-fro drift produced by the vibrations of sound. That diffusion does occur markedly at the jet is shown more strikingly by streams of burning coal-gas or of hydrogen. Any jet of burning coal-gas exhibits an external luminous boundary which is greater than that of the orifice from which it issues, and the amount of spreading at the orifice increases as the velocity of the stream is reduced. The magnitude of this diffusion at the jet was well shown by an experiment in which burning coal-gas issued from a hole 1 mm. in diameter in a brass plate: when the pressure was reduced to a very low value, so that the flame was almost hemispherical, the diameter of the luminous outer boundary at the surface of the plate was 5 mm., and when hydrogen was substituted, the diameter was 8 mm. The increase in the case of hydrogen is to be expected in view of the fact that the hydrogen in the coal-gas is mixed with slow-moving molecules and its diffusion velocity is consequently less. This spreading at the jet is just observable in air streams and was estimated to be of the order of one or two-tenths of a millimetre. In figure 4 (*A*–*C*) plate 1, for instance, light reflected from the slightly chamfered edges of the slit shows *through* the smoke.

Diffusion does, of course, take place at every point of the boundary of the stream and is not restricted to the orifice, but it is only at, or very near, the orifice that the molecules are approximately at rest. Further along the stream the velocity of the central portions of the stream has been communicated to them owing to the fact that the gas possesses viscosity; and moreover this is handed to the external air which starts to move in the same direction. In these regions, then, the lateral diffusion-velocity has to be compounded with the much greater velocity of the edges of the stream, and the resultant motion is only inclined at a very small angle to the axis of the stream; cf. figure 5*g*, plate 2.

The diffusion at the jet is no doubt further favoured by the stagnant region which in most cases exists in its neighbourhood owing to the fact that the circulation of the external air is such as to involve a turn through a right-angle at this point. This is illustrated diagrammatically at *a* in figure 11 (*A*). At first sight, it might seem that impaired sensitivity would ensue if the jet had a razor edge, and to test this a thin circular brass jet sharpened to a fine edge was compared with a circular glass tube of the same diameter with the usual blunt end. No difference could be noted and it was found that the stream spread downwards over the edge, this process being much more marked in the case of burning coal-gas, and this is illustrated in figure 11 (*B*). A stagnant area exists also in this instance, and it was found by moving a shield about that this was the region in which the stream was most sensitive.

Perhaps it would be well at this point to mention an apparent exception to our view that the jet is an essential factor. This exception is a flame such as that of the candle, and this, as we should expect, is insensitive in the ordinary state, but, according to Tyndall[(8)], it can be made to show sensitivity if the flame is distorted by a stream of air impinging on it from a blowpipe. The writer was not able to repeat this, but from the consideration which has just been put forward it seems very unlikely that conditions for sensitivity can be produced by blowing air at the upper portions of a flame, and a more probable explanation would be that the blowpipe orifice acted as a sensitive jet and the candle flame was only an indicator of the disturbance produced. Flames employed as indicators and radiators of disturbances impressed on them in this manner have been examined exhaustively by Chichester Bell[(9)].

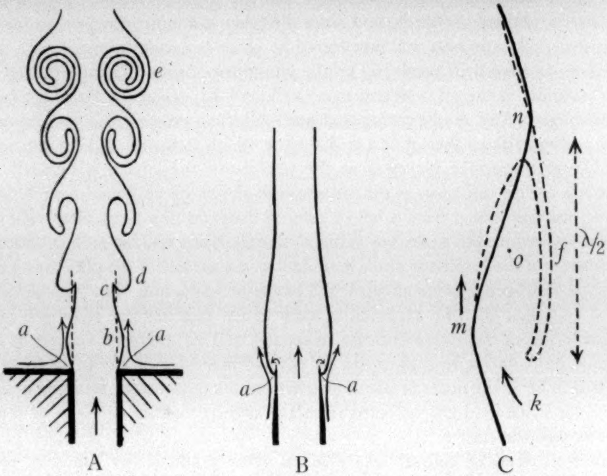


Figure 11.

§ 17. THE EFFECT OF SOUND ON THE DIFFUSION AT THE ORIFICE

We must now consider the effect of sound in the stagnant region near the jet, where there is very little vertical motion and movement is due almost entirely to diffusion. Such movement will be very easily increased or decreased by any drifting of the air as a whole: consequently when a sound-wave approaches the jet, the vibration of the air towards and away from the jet will cause the diffusing particles to advance and retreat, and thus will produce periodic undulations in the boundary of the stream. At the orifice the crest (*b*), figure 11 (*A*), is further forward and the trough (*c*) further back than the normal boundary, but measurement of photographs and careful observations indicated that near the orifice the bottom of the troughs never lay appreciably within an imaginary surface made by projecting the orifice vertically (the inner dotted line). At points removed from the orifice, however, where marked vorticity may be developed, it is possible for the troughs to be well within this surface; cf. figure 4 (*C*), plate 1. This is what we should expect if, in the immediate neighbourhood of the orifice, the stream was surrounded by a thin sheath of slow-moving molecules. The amplitude of these undulations varies with the amplitude of the sound, and an experiment was made to see if the two amplitudes were the same. A tuning-fork was lightly struck and held with one prong close to the jet: by means of the stroboscope a slow-motion view was obtained in which the amplitude of vibration of the prong and of the undulation could be compared. They were found to be the same, and although the experiment was only a rough one it indicates that the amplitude of the undulations is, as we should expect for wide jets, approximately the same as the vibratory motion in the sound wave.

§ 18. VORTEX GROWTH AND DISRUPTION

The fact that shielding the stream of gas from the sound has absolutely no effect, except when the shield covers the first millimetre or so nearest to the orifice, shows that the vibrations of sound play no part in the subsequent development of the undulations produced there. This development must therefore arise through the contact of the crests and troughs with the external air as they move up the column. Observation shows that the first effect of this motion through the relatively stationary external air is to produce in the crests, as we should expect, a flattening in front and an overlapping of the trough at the rear, so that a pear-shaped protuberance results; cf. figure 4 (*B*) and (*C*) and figure 5 (*G*). This is shown diagrammatically at (*d*) in figure 11 (*A*). Further motion causes entrainment of air in the trough and definite circulation then commences; it increases until the vortex occupies half the width of the stream. At this point it comes in contact with the circulation due to the vortex symmetrically placed on the other side of the stream, see (*e*), figure 11 (*A*). The vortex with greatest circulation then forces the weaker one ahead, as it should do even in classical hydrodynamics. This action continues until the vortices are arranged alternately, and they are then free for further development since they can grow until their circulation includes the whole stream. Whether they do so or not depends on the velocity: if the velocity is low, as in figure 5 (*G*), they may become diffuse and tend to disappear; if the velocity is a little higher, as in figure 5 (*F*), they may remain in the same state growing but little; further increase in velocity causes them to grow and pack into one another in the very beautiful manner shown in figure 5 (*E*), and in this stage the vortices at a certain height above the orifice fill the whole of a trough and further entrainment of air therefore ceases. Velocities in excess of this cause the vortices to burst as soon as their circulation involves the whole of the column: their filaments then spread radially, and this motion, combined with the upward motion, produces the well-known wedge-shaped spreading of gaseous streams. Perhaps, then, figure 5 (*E*) can best be described by saying that the vortices are bursting in an orderly manner.

If we are correct in our view that further vortex-development is due to the contact between the undulations produced very near the orifice and the nearly stationary external air, then we should expect that if the troughs and crests are large it will not need so great a velocity relative to the external air to produce overlapping and vorticity, as it would require if they were small. In other words, we should expect the minimum velocity for vortex-production to be a function of the amplitude of the sound wave. An experiment was made to test this, a 1-mm. slit and the stroboscope being used so as to show up disturbances which are invisible to the unaided eye. The results, for a frequency of 100 c./sec., are shown in figure 12. The amplitude was reduced to about one-fifth of its full value and it can be seen that this necessitated doubling the minimum velocity.

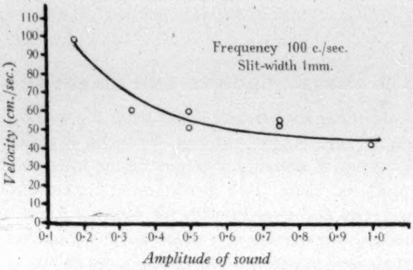


Figure 12. Velocity for minimum effect against amplitude of sound.

§ 19. LIMITS GOVERNING SENSITIVITY TO SOUND

We can now see clearly why there is a minimum velocity, below which gaseous streams are insensitive to sound: the velocity must be great enough to develop the undulation produced at the jet. Further we can also see why, when this velocity is attained, there is then a minimum frequency to which it is sensitive: very low notes will have their troughs and crest so far apart in the stream that only the mildest undulations will be produced even though their amplitude be very great.

No clear limit to the maximum velocity exists, for some change can usually be observed when sound impinges on jets, even when in a state of great turbulence.

Maximum and minimum values of λ, the distance between successive vortices on the same side of the stream, were found in this research to be of the order of 15 mm. and 1 mm. respectively, the maxima occurring in the widest jet and the minima in narrow jets.

§ 20. THE EFFECT PRODUCED BY STATIONARY WAVES

It has so far been assumed that the effect of sound on the stream near the jet is to produce symmetrical swellings and contractions such as are shown in figure 4, plate 1 and figure 5*g*, plate 2. But it is obvious that this cannot be the general case because if it were all jets should be sensitive at the nodes when placed in stationary waves. The air would have to be moving outwards on both sides of the jet to produce the swellings and moving inwards on both sides to produce the contractions. But it is a well-known fact that sensitive flames are most affected at the antinodes.

Repeated attempts were made to demonstrate sensitivity at nodes when the vortices were in the symmetrical position, but in order to get definite stationary waves the loud-speaker and reflecting boards must be large compared with the wave-length of the sound, and unfortunately this condition could not be realized in the apparatus available. However, it was found possible to demonstrate the effect with a jet of width 3 mm. and a frequency of 252 c./sec., giving a wave-length of about 136 cm. At a distance of 68 cm. from the reflector the vortex development was markedly greater than at a distance of 34 cm., showing that the nodes are more effective in this case. With full intensity of the sound some vortex development occurred in any position, but if the intensity was reduced so that the vortices were barely visible stroboscopically at the node, then they disappeared entirely at the antinodes.

To account for the fact that maximum sensitivity is usually shown at antinodes, we must suppose that in narrow jets with consequently high velocities (and these are what are generally employed) the circulations of the two initial vortices interact on one another within a very small distance from the orifice, causing the weaker vortex to be displaced ahead, and that a point arrives at which it is actually easier to form the vortices already displaced, i.e. alternately, than to form them symmetrically.

§ 21. THE CAUSE AND NATURE OF THE VORTICITY

We must now consider the method by which the vorticity in the stream is initiated, and attempt also to explain why, when the vortices are produced, their rate of revolution is proportional to the frequency of the sound.

The first effect of the sound is to produce, as we have seen, undulations in the boundary of the stream and these crests and troughs are carried upwards and have to pass through the relatively stationary air surrounding the jet. This causes the crests to become flattened in front and to overhang the trough following them: thus a filament extends backwards from the tops of the crests. In narrow jets the vortices have reached the alternate position before this stage is reached, and then the whole stream becomes sinuous: the mechanism of the filament-production is, however, similar in both cases. This is shown in greater detail in figure 11 (*C*). One crest and one trough, together making a distance which has been called in this paper one wave-length λ, are drawn with a thick line. The distortion of this formation caused by its motion upwards past the stationary air is observed to take the form shown by the dotted line. Air is entrained in the direction shown by the arrow at *k*, and appears to prevent the filament *f* from extending so as to cover the trough completely. Observation shows that contact of the particles at *n* with the external air reduces their velocity practically to zero.

We have, then, a flattened vortex symmetrical about the point *O* and extending for a distance of about half a wave-length in the vertical direction. Further development consists in the end of the filament turning inwards and then upwards and curling round on the space surrounding *O*, which enlarges to allow of this motion taking place. Now, examination of the photographs shows that the leading filament describes a nearly circular path of constant diameter, and this, of course, must be the case since there is no sink. Consequently the increasing convolutions of the spiral can only be added externally, so that the spiral must grow in size continually.

What relationship now holds between the velocities of different portions of the vortex? Observation shows that the leading filament describes its constant path with the same velocity, so that ω is constant. Owing to viscosity there will be a tendency for the velocity of the portions following and surrounding it to increase, and this does seem to be the case since, during the *second* revolution of the leading filament, the outer filament moves down occupying the whole of the trough and finally making contact with the following crest. That while it does so its velocity with respect to the external air is no longer zero, seems to be indicated by the thin film of smoke which appears to be torn from it and, travelling upwards, produces the straight-line boundary which is such a marked feature of the vortex-development beyond the point where entrainment of the external air ceases; cf. figure 5 (*A*–*E*), and (*a*–*e*), plate 2. This could only occur if there were relative movement between the vortex filament and the external air.

As soon as the descending filament makes contact with the following crest, entrainment of air ceases and the outer filaments thus begin to unite, and from this point onwards the circulation appears to slow down. It seems therefore that the vortex filaments do not have long in which to increase their velocity, and so this probably remains very nearly constant in the general case. A special case arises, however, in streams such as those illustrated in figure 5 (*A*–*C*), in which the velocity is much above its critical value for turbulence and a frequency is forced upon them near the lower limit of their frequency range. The vortices then grow to a very large size and the outer filaments have very much increased velocity, which is shown by the abnormal angular spreading on disruption and by the very marked film which is torn from them as they move downwards towards the jet. In fact, it will be observed that flying filaments of the disrupted vortex may even travel backwards sufficiently far to be re-entrained by a following vortex.

We may now attempt to explain why the rate of rotation is proportional to the frequency of the sound. Let the mean upward velocity of the particles in a thin layer near the boundary at *m* be *V*. The velocity of the particles in *f*, we have seen, is practically zero. Now this incipient vortex is symmetrical about the point *O* which will appear to move upwards with the velocity *u* equal to $V/2$, and relatively to this centre there is an upward velocity of $V/2$ on one side and a downward

V

velocity of $V/2$ on the other. The time of rotation will therefore be the time taken by the filament to travel with a velocity $V/2$ down a distance approximately $\lambda/2$ and up a similar distance, the time taken in turning at the narrow ends being neglected. Hence the time T taken to complete one revolution will be given by

$$T = \frac{2\times\lambda/2}{V/2} = 2\lambda/V.$$

Therefore the angular velocity $\omega = \dfrac{2\pi}{2\lambda/V} = \dfrac{\pi V}{\lambda}$.

But the wave-length is the distance travelled by this thin layer of particles of velocity V in the time of one complete vibration of sound, so that

$$\lambda = V/n.$$

Hence we have $$\omega = \pi n,$$

which is the relation (6) obtained experimentally. Although this treatment is only approximate it does give valid physical grounds for inferring that the rate of rotation of the vortices will be proportional to the frequency of the sound.

We have here taken u, the vortex-velocity, to be half that of the mean velocity V of the stream in a thin outer layer of thickness similar to that of the leading filament. Now as the vortex grows its circulation will enclose faster-moving layers of the stream and this will tend to increase V. But this will only be true up to the centre of the jet: from this point onwards the circulation will include slower- and slower-moving layers until finally the whole of the stream is involved in the circulation of one vortex. We should therefore expect the relative vortex-velocity to increase up to the point at which the circulation includes half the stream, and to fall off after this point has been reached, and this is just what we have seen in the graph in figure 8: u/U rises until the vortices adopt the alternate position (i.e. occupy half the stream) and then falls off to what appears to be a constant value when the circulation involves the whole of the stream.

§ 22. THE INSTABILITY OF GASEOUS JETS IN GENERAL

The conclusions to which we have been led during the course of our consideration of the phenomena exhibited by acoustically sensitive jets allow of a certain amount of light being thrown on the question of the instability of jets in general, and especially on the cause of the wedge-shaped or conical spreading of jets which, although so familiar, has not yet received an adequate physical explanation.

When a gas first escapes through a jet it develops a mushroom-shaped head owing to its action in pushing aside the stationary gas outside, and this causes a circulation in the outer portions producing what is known as Overbeck's vortex. The latter moves upward and is followed by a column of gas whose sides are inclined at a very small angle outwards, representing the resultant motion of particles having an upward velocity together with a lateral velocity due to diffusion; cf. figure 5 (*g*). In the upper portions the velocity of the stream falls gradually to zero owing to the transfer of velocity to the outside air by virtue of viscosity; the boundary then becomes more and more diffuse. At this stage, disturbances of large amplitude may affect the sensitive region near the jet, but any vortex motion which may ensue soon dies away and is not visible to the unaided eye. As the velocity is increased it approaches the critical value for sensitivity; and disturbance then produces vortices which travel up the column and cause the boundary of the stream to widen; cf. figure 5 (*g*). This is the cause of the flickering appearance presented to the naked eye.

Increase of velocity beyond this critical value initiates what we may call the *sound-maintained turbulence stage*, and the stream then adopts, independently of any outside disturbance, the well-known form of a column which, at a certain height, spreads into a cone in the case of circular jets, or into a wedge when the orifice is a slit. In this stage the velocity is sufficient to develop the vortices so that they reach the alternate stage and then increase until their circulation occupies the whole width of the column, at which point, as we have seen, they disintegrate. By listening through a tube applied to different parts of the stream it can be found that the sound is produced in the region where the air is entrained and is generally audible as a low hissing noise: in ignited streams it produces what is usually termed a "roar". The turbulence in the stream is now self-maintained and very irregular, but it can be forced by sound of a definite pitch and then presents a periodic appearance which we have noted for instance in figure 5 (*A*–*C*).

Further, if we remember the results of the investigation discussed in § 5 and illustrated in figure 7, viz. the approximately linear relation between the velocity of the stream and the lower limit of frequency-sensitivity, we see that as the velocity is increased the frequencies affecting it become higher and so if the turbulence is to be sound-maintained the pitch of the sound emitted will have to rise with the velocity. That the general tone emitted by turbulent jets does rise with the velocity, and that the relation is a linear one was first observed by Kohlrausch(10).

The radial expansion of the vortices due to centrifugal force combined with their upward velocity is the cause of the wedge-shaped and conical spreading of gaseous jets. The amount of this irregular spreading will depend in a complicated manner upon the ratio of the lateral velocity of the rotating filaments, at the moment when the circulation of the vortex occupies the whole of the stream, to that of the vortex itself, and upon the motion of the external air. Now we have found the angular spreading to be dependent only on the value of u/U and that when this ratio is constant there is a linear relation between n and U (§ 8). But this is just what was found by Kohlrausch for the general tone emitted and the velocity. On the assumption that the turbulence is maintained by the sound it produces this indicates that the angle of the wedge or cone will not increase beyond a certain amount but will remain sensibly constant, a characteristic of turbulent jets first recorded by Thomas Young(11).

Finally, when the Reynolds critical velocity is reached the stream emerges already in a state of turbulence, and the apex of the wedge or cone then jumps suddenly to the orifice itself, where it remains for all subsequent increases in the velocity.

§ 23. SUMMARY AND CONCLUSION

The secret of the acoustical sensitivity is found to lie in the fact that the portions of the stream in contact with the solid boundary of the jet emerge with very small velocity and spread laterally owing to diffusion. This lateral drift is very sensitive to the to-and-fro motion of the air caused by the vibrations of sound, and the result is that undulations are produced in the boundary of the stream which, on being carried upwards, develop into vortices owing to the friction between them and the stationary air through which they travel. A solid jet from which the gas issues is therefore an essential feature of the sensitivity, and so we must "look upon all jets as musically-inclined", in the words of John Leconte who discovered sensitive flames.

The subsequent growth of the vortices entrains air which, in ignited streams, increases the combustion and thus causes a marked drop in height, the characteristic phenomenon to which sensitive flames owe their discovery and their use.

The most surprising property of sensitive jets, viz. that certain frequencies are more effective than others in producing vortex development whatever the velocity of the stream, the size of the jet and the amplitude of the sound, is found to be connected with the rate of revolution of the vortex, which is the only feature of the vortex motion that depends only on the frequency.

With the knowledge gained of the vortex motion produced by pure notes it has been found possible to throw some light on the physical processes underlying the instability of gaseous jets in general. It has been found that three stages occur: firstly, a non-turbulent flow; secondly, a turbulent flow in which the turbulence is due to the sensitivity of the jet to sounds produced by itself; and thirdly, a stage in which the turbulence already exists in the jet before it emerges and is due to the velocity being above the Reynolds critical value. In particular, it has been found possible to account in a general way for the constancy of the angle of the wedge or cone into which the column spreads during the sound-maintained turbulence stage. It appears that the reverse of Tyndall's explanation of sensitive flames is really the case. Tyndall said that jets were sensitive to sound because they were readily turbulent; actually they are readily turbulent because they are sensitive to sound.

Every effort has been made to treat the matter as fully as possible, in the hope that mathematicians may feel induced to address themselves to it, and for the same reason all the reproductions from the cinematograph films have been enlarged to natural size.

It is hoped, in a further research, to apply the same treatment to the elucidation of the vortex motion produced by edge tones, the mechanism of which is very similar to that of sensitive jets.

§ 24. ACKNOWLEDGMENTS

In conclusion, the author desires to express his gratitude to Prof. E. N. da C. Andrade, in whose laboratory this research was done, not only for providing every facility for it, but also for his original suggestion that the subject of sensitive flames might be one which would repay a searching investigation.

REFERENCES

(1) Brown, G. B. "On Sensitive Flames", *Phil. Mag.* **13**, 161 (Jan. 1932).
(2) Brown, G. B. *Science Progress*, **104**, 672 (April, 1932).
(3) Lewis, J. W. *Proc. roy. Soc.* A, **117**, 393 (1928).
(4) Rayleigh. *Scientific Papers*, **1**, 368.
(5) Richardson, E. G. "Edge tones", *Proc. phys. Soc.* **43**, 239 (1931).
(6) Tolmien. *Z. angew. Math. Mech.* **6**, 468 (1926).
(7) Prandtl. *Z. angew. Math. Mech.* **5**, 136 (1925).
(8) Tyndall, J. *Phil. Mag.* **33**, 92 and 375 (1867).
(9) Bell, Chichester. *Philos. Trans.* **2**, 383–422 (1886).
(10) Kohlrausch, W. *Ann. Phys.*, Lpz., **13**, 545 (1881).
(11) Young, Thomas. *Philos. Trans.* 112 (1800).

DISCUSSION

Dr E. G. Richardson. The clarity of the author's photographs has enabled him to made deductions about the vortex motion in jets which should provide a basis for future mathematicians to attack the theory not merely of sensitive jets but of the efflux of fluids from orifices in general. It is unfortunate that at present the mathematicians, even those trained in the Prandtl school, can scarcely cope with this difficult problem. To my mind the most important observation which the author has made concerns the rate of rotation of the leading filament of the vortices, a result which he deduces (p. 729) theoretically on the assumption that the centre of the vortex moves at half the jet-speed. But does not the assumption imply that no air is entrained by the outer edges of the jet, whereas on p. 715 he adduces entrainment of the air as a reason for the failure of Rayleigh's theory in the case of his jets?

Author's reply: In my attempt to show that the rate of rotation of the leading filament might be expected to be independent of everything except the frequency of the sound, I assumed that the vortex growth commences in "a thin outer layer of thickness similar to that of the leading filament" and that the vortex travels with half the mean velocity of this layer—not half the mean velocity of the stream itself, an assumption which, as Dr Richardson suggests, is not borne out by the experimental results. The treatment would apply, for instance, to vortex growth in the boundary of a semi-infinite stream of fluid. The width of this layer is a function of the amplitude of the sound—cf. figure 5 (*a*–*g*)—and of the width of the stream.

548.56:546.811 548:531.743

THE TWINNING OF SINGLE CRYSTALS OF TIN

By BRUCE CHALMERS, B.Sc., Ph.D., Lecturer in Mathematics and Physics, The Sir John Cass Technical Institute

Received February 14, 1935. *Read May* 17, 1935

ABSTRACT. Experimental methods are described for the production of single crystals of tin and for the determination of their orientations by optical measurements. The conditions under which parts of the crystals can be caused to twin by impact or by tension are investigated, and the determination of the energy relations for certain controlled cases is described. It is found that when twinning occurs, energy amounting to 8×10^5 ergs is converted into heat per cm^3 of crystal twinned. The process of twinning in relation to the crystal structure of tin is discussed.

§ 1. INTRODUCTION

The behaviour of the crystals constituting a metal test-piece under stresses large enough to cause permanent distortion of the test-piece has been the subject of many investigations. It has long been known that one of the ways in which yield occurs is by the glide of lamellae over one another, and experimental work on specimens consisting of single crystals of metals has shown that in many cases the whole distortion can be attributed to such glide(1, 2). It has been shown, however, by S. W. J. Smith, Dee and Young(3), that the formation of Neumann bands in α iron is due to twinning of the volume of which the Neumann band is the trace on the surface of the specimen. The process described is a definite movement of each layer of atoms relatively to the neighbouring layers, so that the atoms then occupy positions in a lattice inclined in a definite direction, that of a twin crystal, to the lattice of the original crystal. The paper cited above gives an adequate account of the geometry of the process of mechanical twinning, but no experimental investigation has been made of the mechanical conditions governing twinning, or of the quantitative energy considerations involved(4).

In order to investigate such conditions, it is necessary to utilize specimens consisting of single crystals, in order that effects occurring inside the volume of a crystal may not be modified by restraints imposed by neighbouring crystals and by inter-crystalline materials. The material with which the present investigation is concerned, namely tin, is particularly suitable, both on account of the ease with which large single crystals can be produced and because it has a crystal lattice which appears to be particularly susceptible to twinning by shock. Work on other aspects of the mechanical properties of single crystals of tin is described and referred to by Obinata and Schmid(5).

The experimental work described below consists in the preparation of the crystals, the measurement of their orientation by a new optical method, a qualitative investigation of the conditions in which twinning occurs, and a quantitative examination of the energy involved in the process.

§ 2. PREPARATION OF THE CRYSTALS

The crystals were prepared by the following method from Chempur tin comprising tin 99·987 per cent, copper 0·00132 per cent, antimony 0·00118 per cent, lead 0·00585 per cent, iron 0·00055 per cent, bismuth 0·00352 per cent, arsenic 0·00005 per cent, nickel 0·00003 per cent, silver 0·00018 per cent, but no zinc, cobalt or sulphur. The tin was melted in a crucible over a bunsen burner and maintained at a temperature of about 300° C., the temperature being measured by means of a thermocouple. Into this was lowered a glass tube of which the bore was the required diameter of the crystal, drawn out at its centre to a narrow capillary. The length of the tube was such that the surface of the liquid tin came up to the bottom of the capillary. When the tube was lowered into the tin it was closed at the top by means of a rubber tube and a clip so that no tin could enter the tube at the lower (open) end. After the tube had been immersed in the tin for a few minutes the clip was opened, allowing the tin to enter, and suction was applied to the rubber tube to draw the tin up the capillary until it solidified at the top. The glass tube was then raised out of the molten tin, the tin inside it solidifying progressively as a rod as it cooled. The tube was raised by means of a lever to which it was clamped, the lever being moved by a clock mechanism which could be adjusted to give any desired speed of raising. It was found that when the rate was of the order of 0·5 cm. per minute a single crystal of diameter 5 mm. and length 6 cm. could usually be obtained. With somewhat slower raising, crystals up to 1 cm. in diameter were obtained.

It was found that when the crystals had cooled they could be made to slide out of the tube without suffering any damage, i.e. without showing either slip bands or Neumann bands. This is due to the fact that tin contracts by about 2½ per cent on solidification. The surfaces of the crystals so prepared were sometimes pitted, but when the tube had been allowed to heat up sufficiently before the tin was admitted, this pitting was usually very slight and sometimes apparently absent. After a crucible full of tin had been in use for some time and had been melted and cooled a number of times it was found that single crystals could no longer be obtained, crystal boundaries being always seen in such specimens. This was taken to be due to impurities in the melt.

Crystals with flat faces were also obtained by inserting slips of mica diametrically in the tubes, so that the tubes were divided into two semi-cylinders. A spectroscopic comparison of the composition of the crystals and of the original tin showed that the amounts of bismuth and lead were slightly reduced by crystallization; the other impurities showed no reduction.

§ 3. DETERMINATION OF THE ORIENTATION

The first step in the investigation of any of the properties of single metallic crystals is the determination of the orientations of the crystal-axes to the geometrical axes of the specimens. The basic method by which this can be done is by the use of X rays, but the process is complicated and laborious. Various alternative methods have been described from time to time, some depending on measurements on the specimen after it has been distorted(6, 7), and some being based on the appearance of the surface after it has been etched(8, 9). The method used in the present investigation is of the latter type, modified so that the orientations can be read off directly and quickly.

The specimen, having been cooled and removed from the tube in which it was prepared, was next etched to prepare its surface for optical examination. It was found that a 5-per-cent aqueous solution of ferric chloride was a suitable etchant for tin. A preliminary visual inspection of the specimen served to reveal any crystal boundaries that were present, and if the whole specimen, or a large enough part of it, consisted of a single crystal, the following property was utilized for the determination of the orientation. It was found that when a slightly converging beam of light was allowed to fall on the surface of the specimen, which was cylindrical in shape, the reflected light did not travel in the directions normally followed when light is reflected from a cylindrical surface. It was instead reflected in a number of definite directions inclined at angles to the geometrical normals of the cylinder.

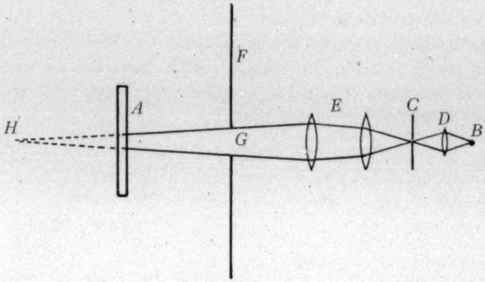


Figure 1.

The directions in which the reflected light travelled were investigated by measuring the positions of the spots of light formed on a suitably placed screen. The optical system employed was that represented in figure 1, in which *B* is a straight filament lamp, the line of the filament being perpendicular to the plane of the diagram, *C* a slit, on to which light from *B* is concentrated by the condenser lens *D*; *E* is a system of lenses forming an image of the slit at *H*. A Microid microprojector is used to give this optical system. The specimen, placed at *A*, is mounted on the rotating axis of a goniometer, while a white screen F with a hole cut in it to allow the beam of light to pass is arranged to receive the light reflected from the specimen.

The light received on the screen consists of a number of well-defined spots, some of which are joined by faint streaks. The number of spots observed while the specimen is completely rotated about its axis depends on the orientation of the crystal-axes and the state of the surface of the specimen. In the most favourable cases as many as 20 definite spots are obtained.

The positions of the spots are measured in terms of two angles, namely the setting of the goniometer when the spot is on a line through G normal to the plane of the diagram, and the angular displacement of the spot from the plane of the diagram along this line. The spots are formed by reflection from planes whose normals do not coincide with the geometrical normal to the surface, so it is reasonable to suppose that the faces concerned are the elementary crystal facets of which the apparently cylindrical surface must in reality be composed. In order to verify this, the angles between the faces giving rise to the various spots were calculated. The tin crystal is known to have a tetragonal lattice in which $c/a = 0{\cdot}541$, and from this ratio the angles between the various simple faces can be calculated. The identification of the angles obtained in a typical experiment with those calculated from the lattice constants is shown in table 1.

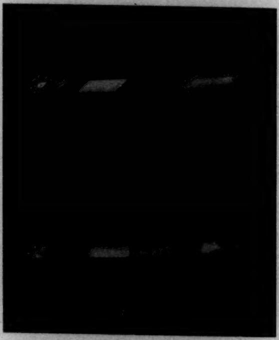
Table 1

	101	001	011	$\bar{1}01$	110
001	27° (28½°)				
011	40° (40°)	28° (28½°)			
$\bar{1}01$	56° (57°)	30° (28½°)	40° (40°)		
110	72° (69½°)	89° (90°)	70° (69½°)	67° (69½°)	
301	28° (30°)	56° (58½°)	61° (63°)	84° (85½°)	50° (53°)

The left-hand figure in each space shows the experimental value of the angle between the reflection normals associated with the two spots concerned, while the right-hand figure, in brackets, gives the value of the angle calculated from the lattice constants. The six spots are referred to by their crystal planes, which were found by a method described below. The agreement between the two sets of figures is such as to preclude any possibility of error in the identification of any spot. The spots observed in various experiments have all been identified as 001, 100, 010; 101, 011, 110; 301, 031; 111, or the corresponding faces with the signs changed. It was found further that the faint streaks joining the spots were only present where one of the three indices was different for the two spots concerned.

The definite identifications of the spots in this manner made it possible to pick out by inspection any required spot from the characteristic pattern on the screen. In particular, the 001 spot is at the centre of a square of the 011 spots, to each

PHYSICAL SOC. VOL. 47, PT. 4 (CHALMERS)



To face page 737

of which it is joined by lines; the distance of these spots from the central spot is fixed for any given distance of the screen from the specimen, and a circle marked on the screen facilitates the identification of a part of this arrangement when it is not all visible. With the visual identification of the 001, or any other, spot, the measurement of its orientation becomes very simple. A holder is fitted so that the specimen can be placed at a definite distance (5 cm.) from the screen, with its axis parallel to the line through G, perpendicular to the plane of the diagram. The specimen is rotated until the required spot is on this line, and the angle between the corresponding normal and the axis of the specimen is read off directly from a scale of angles marked on the line through G. Hence the orientation of any axis to the axis of the specimen can be determined immediately, the accuracy of a determination being estimated as within 1°.

It was found that with specimens prepared in the way described above, in the majority of cases the 001 axis was nearly normal to the axis of the specimen. Specimens were, however, obtained with orientations having various other values.

§4. MECHANICAL TWINNING OF TIN

When certain conditions are satisfied, the effect of the application of a force, either impulsive or steady, to a single crystal of tin is to cause parts of the crystal to become twinned with respect to the original crystal. The conditions under which this happens will first be described generally, and then quantitative results will be given.

Qualitative description. Twinning takes place most readily when an impulsive force is applied to the end of a cylindrical single crystal, the 001 axis of the crystal being roughly perpendicular to the length of the specimen. The twinned portion is always bounded by parallel planes of the 301 type or by the ends of the specimen. Subsequent applications of the same type of force may extend the region twinned or may cause a second part of the crystal to twin; when this occurs, the second part is usually bounded by planes parallel to those of the first part, but occasionally by a second pair of planes of the same type. The plate shows at (*a*) two perpendicular views of a crystal in which two parts between planes parallel to each other are twinned, the darker parts being the parts which have twinned.

The plate shows at (*b*) a crystal on which three different sets of parallel planes have come into operation as twinning planes.

It is possible by continual longitudinal tapping to cause the whole crystal to take up the twin orientation. The twinned part is of the form of an elliptical cylinder obtained by inclining the axis of the original cylinder by about 5° to the normal to the circle forming its base. This angle between the untwinned and twinned parts is visible at (*a*) in the plate.

If the surface of a part of a crystal that has been twinned but not re-etched is examined optically, it is found that the reflection pattern corresponding to the original orientation persists. Etching in ferric chloride solution to remove the surface layer reveals, however, an arrangement of spots corresponding to the orientation of a twin about the 301 plane, which separates the twinned and untwinned parts. A part of a crystal twinned as described above, or a crystal with its 001 axis nearly parallel to the length of the specimen, will not undergo any further change if further longitudinal impact is applied, unless the force is sufficient to cause the specimen to bend with irregular distortion of the lattice. The reason for this is considered in the discussion of results (§ 5).

If, however, a tension is applied to a specimen that has been twinned, it will sometimes revert to its original (untwinned) orientation. This occurs suddenly and a distinct click is audible as any portion of the crystal changes its orientation. The reversion to the original orientation is demonstrated by etching and optical examination.

Reversion to the original lattice can also be brought about by transverse impact in a direction perpendicular to the axis of the specimen in a plane containing the normal to the 301 plane of the original twinning and the axis of the specimen. Twinning on other planes owing to undue violence prevents a succession of more than one or two of the alternate twinnings and untwinnings which should otherwise be possible. Longitudinal compression of a suitable specimen in a vice causes twinning to occur in the same way as by impact, a characteristic click being heard as the twinning takes place.

In general twinning may occur when a compressional force, either impulsive or steady, is applied in a direction perpendicular to the 001 axis, or when a tension is applied parallel to the 001 axis. The limiting variations from these definite orientations within which the twinning occurs have not yet been determined.

Quantitative results. The observations described in the preceding paragraphs indicate that when twinning is brought about by longitudinal impact, the volume of the crystal in which twinning takes place is in some way dependent on the conditions of the impact. This variation was studied by using a ballistic pendulum to apply definite known impulsive forces to the specimen.

Two brass cylinders, each of mass about 200 gm., were suspended with their axes in the same horizontal line by means of a system of threads so that they could only swing in the vertical plane containing their common axis. A hole was drilled to a depth of about 1 cm. in the end of one of these cylinders, and the specimen on which observations were being made was held firmly in contact with the bottom of this hole, and projecting towards the second cylinder, by means of three set screws and a ring of rubber sponge. When both cylinders were at rest the free end of the specimen was just in contact with the end of the second cylinder.

When the second cylinder was displaced from its position of rest by a given amount and released, it struck the end of the specimen, giving to it and its holder an amount of kinetic energy that could be determined from the subsequent swing of the first cylinder. The displacement of each of the cylinders was observed by the movement along a horizontal scale of pointers attached to them. By using a specimen of which the orientation was such that no twinning took place on impact, a calibration curve relating the displacement of the striker and the swing of the holder was obtained; this is the curve *A* of figure 2. When specimens were used in which twinning took place, the points shown in figure 2 were obtained. The horizontal distance between a point and the calibration curve gives some indication of the loss of kinetic energy, and it is apparent that even with the same specimen and the same displacement of the striker, this energy-loss may vary very widely. Hence while some kinetic energy is lost in these collisions, this loss of energy is not directly related to the initial displacement of the striker.

A measurement was also made of the amount of twinning that took place during each collision, the volume twinned being calculated as the product of the cross-section of the specimen and the axial distance between the boundaries of the twinned part. These measurements were made with a travelling microscope.

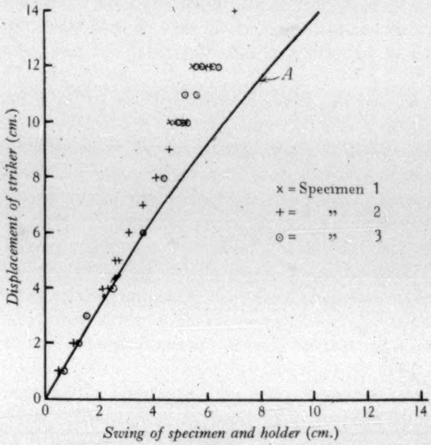


Figure 2.

The calibration curve of figure 2 corresponds to a definite loss of kinetic energy during a collision. A point to the left of the calibration curve corresponds to a greater loss of kinetic energy, and this difference of energy-dissipation was calculated for each collision. A correction, based on a consideration of the momentum, was made for the kinetic energy of the striker after collision.

The points in figure 3 are those obtained when this loss of kinetic energy is plotted against the volume twinned in the same impact. The points, while showing that the accuracy of the observations is not high, indicate that the relation is approximately linear, i.e. that the loss of kinetic energy in an impact which results in twinning is proportional to the volume twinned thereby. The numerical value of this relationship is that the energy dissipated is 8×10^5 ergs per cubic centimetre twinned.

It is noticeable from figure 2 that when the initial displacement of the striker is less than a definite minimum for each specimen no energy loss occurs; this corresponds to collisions in which no twinning takes place. The minimum displacement with which twinning occurs will probably depend on the orientation and cross-section of the specimen; the results obtained so far only indicate that it is greater when the cross-section of the specimen is greater.

If the atoms of the twinned part of a crystal occupy a lattice identical with, but inclined to, the original lattice, it follows that the potential energy of the atoms in the twinned lattice will be the same as that of the atoms in the untwinned lattice; hence none of the kinetic energy that is lost in the collision can be more than temporarily changed into potential energy of the lattice. A small part of the energy will be converted into sound as the characteristic click that always accompanies twinning, while the remainder must take the form of heat.

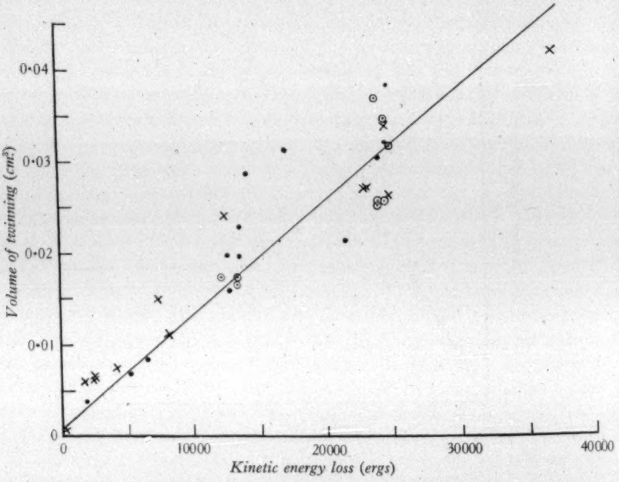


Figure 3.

It follows that the temperature of a part of the lattice that has twinned must be higher immediately after twinning than it was before. The energy-loss given above, 8×10^5 erg/cm.³, would, if completely converted into heat, represent a rise of temperature of 0·05° C. in the twin crystal. This temperature-rise was roughly measured by inserting a thermocouple into a hole drilled transversely in the specimen and observing the deflection of a galvanometer connected to the thermocouple when the specimen was caused to twin by tapping it. The corresponding rise of temperature was found from a calibration of the thermocouple against a thermometer over a wider range of temperature. In a typical experiment the rise of temperature obtained was in two parts, coinciding with two impacts, and amounted to 0·01° and 0·03° C. Further tapping caused no further observable rise of temperature. The fact that the observed rise of temperature is less than the calculated rise is due to the twinning being somewhat irregular as a result of the distortion of the lattice consequent on the drilling of the hole in the specimen.

§ 5. DISCUSSION OF THE RESULTS

In the foregoing sections of this paper it has been assumed that the phenomenon that has been described and measured is a process of twinning. It is necessary now to examine critically the evidence that this is the case.

We may define twin crystals as crystals such that the crystal axes of each form a mirror image of those of the other about the surface separating them. For the present purpose we must identify one of the twin crystals with the original or unchanged specimen, and the other with the part that has been altered by the treatment applied to the crystal.

In the first place it can be shown that the changed part is itself a single crystal; this has been demonstrated in the following ways: (i) after etching, the surface of the changed part can be made by the method of § 3 to show a definite series of reflection spots which are quite different in their disposition from those of the unchanged crystal, and are not altered by further prolonged etching; (ii) when the changed part of the crystal is stretched, in certain cases it yields by glide, with the formation of slip bands; (iii) by the application of a suitable force, either impulsive or steady, the orientation of this part of the specimen can be made to revert to that of the original crystal, as determined by the reflection spots; (iv) that the effect is not confined to the surface layers of the specimen can be shown by polishing and re-etching the specimen; the arrangement of the reflection spots is not altered by this treatment.

The relation between the lattices and the plane separating them must now be considered. The plane of separation is found by calculation from the reflection spots of both parts of the crystal to be within 2° of the position of a plane of 301 type of both lattices. Since such a plane of separation must be a definite low-index crystal plane, it follows that the measurements are sufficiently accurate to show that it is a 301 plane.

A more elegant method of showing the two crystals to be twins about the 301 plane depends on the relation between the reflection spots of the two surfaces. The diagram, figure 4, represents the section of the lattice of tin by the 010 plane, the crosses and dots representing the atoms in alternate layers. The structure is a tetragonal lattice with c/a equal to 0·541, with atoms at the corners of the unit cell and at the centres of the rectangular faces; such a unit cell is indicated in the bottom left-hand corner of the diagram. The sections of the densest planes perpendicular to the plane of the diagram are also shown on the left-hand side of the diagram.

The line AB represents the trace of a 301 plane on the plane of the diagram; the crosses and dots surrounded by circles to the right of this line represent a lattice which is a twin of the original one about this plane. The traces of the more important planes of the twinned crystal are marked on this part of the diagram, their identity being indicated by figures enclosed in round brackets. The planes of the original lattice are indicated by dotted lines in the same part of the diagram. It can readily be seen from the diagram that the following pairs of planes in the twinned and untwinned parts (the round brackets indicating the twinned lattices) are only separated by small angles, the values of which are marked on the diagram: 100, ($\bar{1}$01); $\bar{3}$01, (001); $\bar{1}$01, (101); 001, (301); 101, (100). It follows from the fact that the reflection spots are caused by the reflection of light from these crystal planes that certain spots in the reflection pattern of the twinned lattice will be only slightly displaced from the positions of spots coming from the untwinned lattice; other

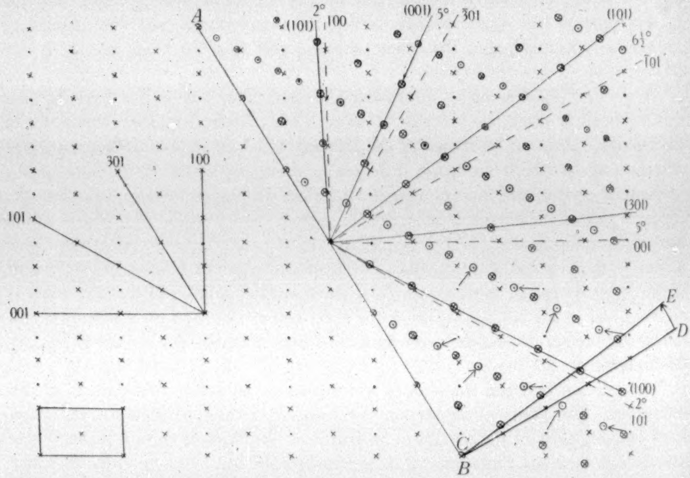


Figure 4.

spots, however, that are present in the untwinned pattern will not appear in the twinned arrangement. The most prominent spots to which this applies are the 011 spots of the untwinned lattice. Figure 5 represents the relative positions of the various spots of which some or all may be seen when reflection takes place from the two parts of the crystal, the same convention as to notation being observed as before. (The figure strictly represents the positions which the spots would occupy on a sphere whose centre is the point of reflection.) It can be seen by inspection of the spots actually observed when reflection takes place from the two parts of the specimen, either consecutively or simultaneously, that the relations between the two lattices are as indicated in figure 4.

In all cases in which these methods were applied, the twinning plane was found to be of the 301 type, of which there are four sets in the crystal lattice. Twinning on three of these planes is clearly visible at (*b*) in the plate.

It is next necessary to consider the movements of the atoms which correspond to the process of twinning, i.e. their movements from their positions in the original lattice to their equally stable positions in the twin lattice. A mechanism suggested by Edwards$^{(10)}$ is obviously contrary to the data described above, and will be disregarded. It is clear from the lattice diagram, figure 4, that the simplest movement of the atoms indicated by crosses is a rotation of each $\bar{1}01$ plane through $6\frac{1}{2}°$ towards the 101 plane (i.e. anticlockwise in the diagram), as indicated by the line *CD* rotating to the position *CE*. Each atom concerned (i.e. three out of every four in the whole lattice) is now in its position in the new lattice. The atoms in the alternate planes cannot take up their new positions in such a simple manner; their probable movement is indicated by the arrows on the lower right-hand corner of the diagram. That the scheme described is correct for the more densely packed planes is confirmed by the observation described above, that the axis of the twinned part is rotated through an angle of about 5° to that of the untwinned part. It can be seen at (*a*) in the plate that the sense of this rotation agrees with that expected from the lattice diagram, since the geometrical axis of the specimen is roughly in the 001 plane of the untwinned lattice.

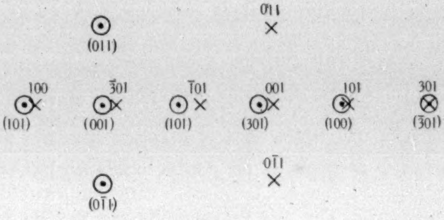


Figure 5.

We can now see why twinning only occurs when certain relations between the directions of the applied force and the crystal axes are satisfied. The lattice diagram shows that the force applied must cause shear to occur in a definite sense between the planes in which twinning occurs. A sufficient impulsive shear stress will give the atoms sufficient kinetic energy to overcome the forces that retain them in their original arrangement, and to allow them to take up their new positions. The kinetic energy will afterwards be dissipated as heat. When the force applied is steady, the elastic displacements of the atoms eventually reach a value such that they are in positions where forces act to bring them into the new lattice.

In either case, there must be a sufficient component of shear stress in the right direction in a set of 301 planes for twinning to occur. When a compressive force is applied perpendicularly to the 001 axis, there may be such components in all four of these sets of planes; the crystal shown at (*b*) in the plate shows regions of twinning on three of these sets of planes. Tension in a single crystal rod perpendicular to the 001 axis, or compression parallel to this axis, will not provide components of shear in the right direction in any of the twinning planes; hence no twinning occurs in either of these cases. When tension is applied in a direction nearly parallel to the 001 axis, the result may be either twinning or glide, depending in some way not yet analysed on the exact orientation.

When there is a component of shear stress in the appropriate direction in two or more of the twinning planes, the plane on which twinning takes place appears to be the one in which this component is the largest, although the numerical results are not yet sufficient to establish this principle definitely.

It has often been observed in the past* that when twinning has been caused by straining in a polycrystalline metal, recrystallization tends to start from regions at which twinning has occurred, and it is suggested that this is due to instability of the twin crystal. This phenomenon has not been observed in the present investigation, although twinned crystals have been kept above the recrystallization temperature for some weeks. An explanation lies in the fact that with single crystals, in which twinning has occurred throughout the whole cross-section, there is no region in which the conditions approximate to those of an intercrystalline boundary, i.e. there is no region of misfit necessitating the presence of atoms not in a lattice. In a polycrystalline specimen, on the other hand, a twinned section of a small crystal will not fit perfectly the neighbouring crystals on which it abuts; further, the change of shape due to the shift of 5° which occurs with twinning will set up a state of strain in a polycrystalline specimen although not in a single crystal. Both these factors explain why a twinned region of a polycrystalline specimen may serve as a nucleus for recrystallization while a twinned single crystal does not.

From the regularity with which the transformation occurs, and the agreement found between various specimens in the present experiments, it would seem that whereas the resistance to glide is a structure-sensitive property, the incidence of twinning is non-sensitive. This may be due to the fact that the process of glide chiefly concerns a few (possibly specially constituted) lattice planes, while twinning concerns every atom in the twinned lattice. Since every atom takes part in the twinning process, it is possible to express the energy relation that results from the ballistic experiments as the mean energy per atom that must be supplied to cause twinning by impact. The figure obtained is 2×10^{-17} erg per atom. The mean energy per atom, however, does not give any exact information, because different atoms play different parts in the twinning process and probably require different amounts of energy. It is clear, however, that the energy is small compared with that which would entirely overcome the cohesion of the lattice and cause melting.

§6. ACKNOWLEDGMENT

In conclusion, the author wishes to record his appreciation of the interest taken in this work by Mr D. J. Macnaughtan, Director of Research of the International Tin Research and Development Council; and of a grant from the International Tin Research and Development Council; and to make acknowledgment to the Governors and Principal of the Sir John Cass Technical Institute for facilities provided. Finally, his thanks are due to Prof. E. N. da C. Andrade, F.R.S., for his continued interest in this work and his many valuable suggestions during its progress.

---

* See, for instance, reference (3).

REFERENCES

(1) A. Ewing and W. Rosenhain. *Philos. Trans.* **193**, 353 (1900).
(2) G. I. Taylor. *Proc. roy. Soc.* **102**, 643 (1922).
(3) S. W. J. Smith, A. A. Dee and J. Young. *Proc. roy. Soc.* **121**, 477 (1928).
(4) E. Schmid. *Internat. Conference on Physics, London,* 1934, Vol. II, p. 168.
(5) J. Obinata and E. Schmid. *Z. Phys.* **82**, 224 (1933).
(6) B. Chalmers. *Phil. Mag.* **14**, 612 (1932).
(7) R. Roscoe and P. J. Hutchings. *Phil. Mag.* **16**, 703 (1933).
(8) P. W. Bridgman. *Proc. Amer. Acad. Arts Sci.* **60**, 305 (1925).
(9) G. Tamman. *J. Inst. Met.* **44**, 29 (1930).
(10) C. A. Edwards. *J. Inst. Met.* **14**, 116 (1915).

DISCUSSION

Prof. B. P. Haigh. The effective demonstration of the mechanical reversibility of the process of twinning and of its thermally irreversible character suggests that it may be an important feature of mechanical strain in other metals as well as in tin. It seems possible that twinning and untwinning may be the direct cause of elastic hysteresis, the nature of which has not yet been explained.

In a paper contributed in 1927 to the Faraday Society* the present writer showed that hysteresis exhibits different characteristics in three distinct stages in the course of fatigue tests on mild steel. In the first and third stages slip bands are produced on the surface of the test piece subjected to cyclic variations of stress; and the plastic hysteresis observed in these stages may be attributed in large measure to the process of slip which probably includes both the production of the amorphous phase and its partial recrystallization with liberation of heat. In the second stage, however, no slip bands are observed, but heat continues to be liberated during long periods which may include many millions of cycles of stress. The characteristics of twinning, demonstrated so clearly by the author, suggest that twinning and untwinning occurring in a cyclic process during successive variations of stress may be a probable cause of this elastic hysteresis as exhibited in the second stage of fatigue tests; and it appears that the subject deserves further investigation with this possibility in mind. It is clear that tin is admirably suited for the investigation of twinning phenomena and that if twinning is not as readily observed in other metals it may be because untwinning occurs more easily on relaxation or reversal of stress.

Prof. E. N. da C. Andrade. The author has accomplished a valuable piece of work in measuring for the first time the energy per cm.^{3} required to produce twinning in a given crystal, and probably owes his success in this difficult task to his ingenious method of utilizing shock to produce the twinning and measure the energy. The obvious course of direct thermal measurement would be very difficult to carry out with the same degree of precision. Many interesting extensions of his work suggest themselves: for instance, how does this energy vary with the temperature of the crystal? One would naturally suppose that less energy would be necessary to produce twinning at higher temperatures, since the heat agitation throws the atoms further from their equilibrium position, but a quantitative estimate of the variation would throw light on the complicated question whether twinning or glide takes place in a given lattice at a particular temperature.

---

* B. P. Haigh, *Trans. Faraday Soc.* **24**, February (1928).

It would be of very great interest for the same question to know what direction of blow, relative to the crystal axes, is most favourable for twinning. Can the author give us any closer estimate than the general indication of § 4 of his paper?

The fact that the whole of the twinned part of the crystal turns through 5° indicates that the modification of the lattice is propagated from the twinning boundary. The readjustment of the atoms into the twinned lattice without any atom moving through more than the interatomic distance is conceivable, if the twinning takes place simultaneously throughout the volume. It might be worth while to find out whether twinning would take place if the crystal were enclosed in an unyielding cylinder and struck, and, if so, whether it would be in close narrow bands alternating in orientation.

I was struck by the use of the optical method for identifying the crystal axes. It has been utilized once or twice previously, notably by Bridgman,* but it seems to have been too much neglected if it is as simple as it appears to be from the author's demonstration.

Mr E. J. Daniels. The author's method of determining orientation appears to be more readily applicable in metallographic investigations than those discussed by Tammann in the paper to which he refers. The observation, recorded on page 739 of the paper, that twinning produced by compression can be reversed by tension giving again an untwinned crystal, seems very illuminating in connexion with a suggestion of Elam† that certain results given by annealing of crystals bent in various manners can only mean that some of the strain set up by bending one way is reversed when the metal is bent in the reverse direction.

Author's reply. Prof. Haigh puts forward a very interesting suggestion regarding the relation between twinning and fatigue; it would seem, however, that a considerable amount of experimental work must be done on metals other than tin before the conclusions arrived at with respect to tin can be regarded as generally applicable to metallic crystals. At the same time, it seems likely that the mechanical reversibility of the twinning process is shared by every metal in which twinning can occur.

I should like to thank Prof. Andrade for the suggestions he has made for future work on twinning. I am not yet in a position to answer any of the specific questions that he raises, but I hope to deal with them in a future communication.

In reply to Mr Daniels, it seems clear that any distortion resulting in the formation of a lattice identical with, but inclined to, the original lattice can be reversed by the application of a force bearing the same relation to the new lattice as the previous force bore to the original lattice. The fact that this takes place as described in the paper indicates that the twinned lattice is identical with, though inclined to, the untwinned lattice.

---

* *Proc. Amer. Acad. Sci.* **60**, 305 (1925).

† C. F. Elam, *Distortion of Metal Crystals*, p. 175.

537·59

AN ATTEMPT TO ANALYSE COSMIC RAYS

By ARTHUR H. COMPTON, University of Chicago
and Oxford University

The Twentieth Guthrie Lecture, delivered February 1, 1935

ABSTRACT. The coincidence experiments of Bothe-Kolhörster, Rossi and others, considered in the light of various alternative interpretations, show the existence of penetrating, electrically charged particles which are either primary cosmic rays or are secondaries of primaries that are absorbed high in the atmosphere. The shower-producing radiation, which seems to consist of photons, must, in view of the marked latitude effect to which it is subject, be produced by some electrically charged primary rays, which are provisionally identified as electrons.

The variation of cosmic-ray-intensity with latitude is shown to increase from about 16 per cent at sea-level to a ratio of 40-fold between 55° and 20° at very high altitudes. Extrapolation to the top of the atmosphere near the poles and the equator would indicate a ratio greater than 100 : 1, which would imply that less than 1 per cent of the unfiltered primary cosmic rays are electrically neutral. In conjunction with the coincidence experiments, this means that *the primary rays responsible for cosmic-ray effects are nearly all electrically charged particles.*

A method of analysing these electrically charged rays is described and applied to typical ionization data for different altitudes at various latitudes. The method consists in (i) calculating the effective minimum energy of various types of particles admitted through the earth's field at different magnetic latitudes; (ii) determining the minimum ranges in air of the particles corresponding to these minimum energies; and (iii) analyzing the data relating the altitude with ionization and thus obtaining experimental range-minima that can be compared with the calculated values. *The balloon experiments show two such range-minima*, A *and* B, at higher and lower altitudes respectively. Their ranges progress with changing latitude approximately in the manner that the theory predicts. *A third range-group* C *appears on analysis of the curve showing the latitude effect for sea-level.* Using the best available information regarding magnetic latitudes and the relation between energy and range, *these range-groups are identified respectively as alpha particles, electrons and protons.* Comparison with directional experiments indicates that the electron group probably consists of about equal parts of positrons and negatrons.

The possible errors involved in applying this analysis are of such magnitude as to leave the identification of the range-groups questionable. Comparison with other data is on the whole however confirmatory, except for the failure to find proton tracks in Wilson photographs of cosmic rays. For this a possible explanation is offered.

§ 1. COINCIDENCE OBSERVATIONS

The first important information regarding the nature of cosmic rays came from Bothe and Kolhörster's famous coincidence experiment, in which two Geiger-Müller counting-tubes$^{(1)}$ were used. The counting-tubes were placed one above the other as shown in figure 1, and were shielded by heavy blocks of lead and iron. Between the tubes, at A, could be inserted a 4·1-cm. block of gold. In their experiment these observers found that the gold reduced the number of coincidences per hour to 0·78 of the value it had without the gold. This reduction is however of the same order of magnitude as the reduction in intensity of the cosmic rays which is observed in ionization chambers when they are shielded by an equivalent screen. Thus Bothe and Kolhörster concluded that the rays that produce the coincidences are the same as the penetrating cosmic rays that affect the ionization chambers. Assuming that coincidences can be produced only by penetrating particles which ionize continuously along their paths, and since the only known particles that produce such continuous ionization are those which are electrically charged, Bothe and Kolhörster concluded that the cosmic rays were some kind of charged particles.

Several alternative interpretations of this experimental result have been proposed: (i) Primary photons may produce secondary β-rays, proceeding in nearly the direction of the primary rays, at such frequent intervals that there is a good chance of ionization occurring within a Geiger-Müller tube placed anywhere in the path of the photon⁽²⁾. Experiments by Street and Johnson⁽³⁾ have shown that the probability of ionization being produced in such a tube by a coincidence-producing particle is at least very nearly 100 per cent. On the other hand, Anderson's cloud-expansion photographs show pairs and showers of high-speed electrons emerging from lead plates into which no ionizing particle enters⁽⁴⁾. He finds it necessary to ascribe each of these showers to a photon, which does not ionize the gas in the chamber until at some cataclysmic event in the lead plate it spends itself in the production of a group of high-energy electrons. The energy of these photons, as determined by summing the energies of the secondaries they produce, is of the order of 10^9 electron-volts, which is within the range of primary cosmic-ray energies. The photographs show however above the lead plate an open region where the photon does not produce ionization. Such photons accordingly do not have ionizing particles associated with them continuously along their paths, as they would need to have if the production of coincidences were to be accounted for in this manner.

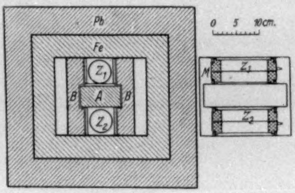


Figure. 1 Experimental arrangement of the two Geiger-Müller counting-tubes used by Bothe and Kolhörster.

A similar conclusion can be drawn from experiments, such as those of Rossi and others, with three counter tubes out of line, in which case a few coincidences are observed which may be ascribed to secondary rays. It is found that by proper shielding of the counter tubes to absorb the secondary rays, such coincidences may be almost completely eliminated, whereas with similar shielding the coincidences occur at a normal rate if the counters are in line. It is thus evident that in the latter case the coincidences are due, not to any secondary rays, but to the direct action of individual penetrating particles.

(ii) Another suggestion has been that the primary cosmic ray is a photon, which however gives rise to a secondary β-ray of penetrating power nearly the same as the primary photon. That this suggestion is untenable has been shown by another coincidence experiment performed in slightly different forms by Rossi[(5)] and in our laboratory by Hsiung[(6)]. Triple coincidences for three counters in line are observed, with lead shields arranged to prevent the counters as far as possible from being affected by secondary rays. A heavy lead block (20 cm. thick in Hsiung's experiment) may be placed, as in figure 2, either between the first and second counters or above the first. If the secondaries have no effect, and if all of the coincidence-producing particles come from above the apparatus, since the fraction of the particles stopped by the lead in cases *B* and *C* of figure 2 should be the same, the coincidence rate in these two cases should be identical. If on the other hand non-ionizing photons

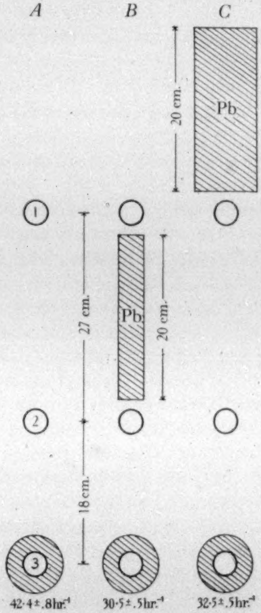


Figure 2. Hsiung's coincidence experiment.

produce secondary penetrating electrons (coincidence particles) in the lead block, the coincidence rate should be greater with arrangement C than with B, since in the latter case these newly-formed particles would not traverse counter 1. After making allowance for the slight effect of the well-known soft secondary rays, both Rossi and Hsiung find the coincidence rate in cases B and C identical within experimental error. This means that all the coincidence-producing particles originate above the apparatus. Thus the proposed interpretation, which would require the production of a considerable number of penetrating secondaries in the lead shield, is untenable.

§ 2. COINCIDENCE RAYS AND SHOWER-PRODUCING RAYS

Rossi(7) has brought forward strong evidence to the effect that the rays which produce the coincidences which we have been discussing are not identical with those which produce the showers of ionizing particles observed in Wilson chambers and with coincidence-counting tubes out of line. Bothe and Kolhörster's early conclusion that the particles producing the coincidences are absorbed by matter at substantially the same rate as the total cosmic-ray beam observed with ionization chambers has been confirmed by later experiments. Thus Hsiung(6) finds 0·016 cm.$^{-1}$ for the absorption of the coincidence particles in lead, and calculates 0·014 cm.$^{-1}$ in lead from the depth-ionization data. Rossi(8) and Johnson(9) find however that with increasing altitude the number of coincidences observed with the arrangement of counters shown in figure 3 (due to showers) increases more rapidly than those with the three counters in line. Thus between sea-level and 2370 metres Rossi and Benedetti(8) find a ratio of 1 : 5 for the coincidences when the counters are out of line and 1 : 2 for those when the counters are in line. Street and Johnson(10) have given direct evidence that some showers result from secondary rays excited by penetrating primary corpuscles, and this seems to be supported also by the evidence obtained by Anderson and Blackett from their Wilson photographs. The difference in the relative number of showers at different altitudes would seem however to require the assumption of an additional shower-producing radiation which is more strongly absorbed in the atmosphere than are the corpuscles responsible for coincidences of the Bothe-Kolhörster type.

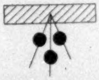


Figure 3. Rossi's arrangement for studying shower-producing radiation.

Experiments with counter tubes and Wilson chambers agree in giving convincing evidence(11) that the direct agent producing the showers of corpuscles responsible for coincidences with the counters out of line is a non-ionizing radiation with a mass-absorption coefficient of about 0·05 gm.$^{-1}$ cm.^{2} in lead, and roughly proportional to the atomic number. In its non-ionizing property and its ability to produce "photo-electronic" pairs of positrons and negatrons, this shower-producing radiation is identical with hard gamma radiation, and almost certainly consists of photons.

§ 3. THE LATITUDE EFFECT

When Bothe and Kolhörster first drew the conclusion from their coincidence experiment that the primary cosmic rays are electrical particles, they remarked(1) that these particles should on approaching the earth from the outside be deflected by the earth's magnetic field so as to reach the poles more easily than the equator. J. Clay had just published(12) his first measurements indicating a greater cosmic-ray-intensity in Holland than in Java. This effect Bothe and Kolhörster ascribed to the anticipated action of the earth's magnetic field. Millikan and Cameron, however, had reported no measurable difference in intensity between California and Bolivia(13), Bothe and Kolhörster(14) were themselves unable to detect a difference between Hamburg and Spitzbergen, as also were Kerr Grant(15) between Melbourne and Antarctica, and Millikan(16) between California and Hudson Bay. In spite of Clay's confirmation(12) of his own findings, it was therefore supposed(17) that the earth's magnetic field had no effect on the primary cosmic rays, and that the effect found by Clay was due to some more obscure cause(18). The absence of such an effect could have been reconciled with the coincidence experiments by supposing that the primary cosmic rays are uncharged particles which are absorbed high in the atmosphere, there giving rise to the penetrating electrical particles which the coincidence experiments show at the earth's surface. Then came, however, an extensive series of observations which established the existence of Clay's latitude effect, and has thrown a flood of new light on the problem of the composition of the primary cosmic rays.

By the work of Störmer(19), Epstein(20), Rossi(21), Lemaitre and Vallarta(22), and others it has been shown that for electrons of a definite energy approaching the earth isotropically from remote space, at magnetic latitudes less than a definite limit the electrons are unable to reach the earth (Störmer), at latitudes greater than a second limit the number of rays reaching the earth is unaffected by the earth's magnetic field (Lemaitre and Vallarta), and at intermediate latitudes the rays can reach the earth from some directions but not from others. For electrons of energy more than 2×10^{10} e.V., the effect of the earth's magnetic field is negligible at all altitudes. Electrons just able to traverse the earth's atmosphere must have an energy of about 6×10^9 e.V., for which the upper and lower latitude limits are 45° and 35° respectively. In figure 4 is shown a family of very useful curves, calculated approximately by Lemaitre and Vallarta, which describe the intensity of the electrical rays at different latitudes for particles with different energies. It is clear from these curves that if the incident particles have a wide range of energies, there should be observed a gradual increase in intensity from the equator toward the poles. Since the earth's atmosphere does not permit particles with less energy than about $2{\cdot}3 \times 10^9$ e.V. (as calculated for protons) to reach sea-level, measurements made at sea-level should show such variations only up to about 50° magnetic latitude. At higher altitudes, where lower-energy particles are transmitted, the latitude effect should extend to correspondingly higher latitudes.

The recent cosmic-ray surveys have given results in complete accord with these calculations. During 1931 to 1933 we had twelve different expeditions, manned by some eighty cooperating physicists, making measurements at more than one hundred stations widely distributed over the earth. The data showed(23) that at north and south magnetic latitudes higher than 50° no significant variation with latitude occurs at sea-level. From the equator to 50°, however, there is at sea-level an increase of intensity of about 16 per cent. Similar contemporaneous measurements by many independent observers have led to essentially the same results. In figures 5 and 12 are collected typical data(23, 24) showing the intensity of the cosmic rays at sea-level as a function of the geomagnetic latitude.

There are several interesting features of these data which cannot be elaborated here. Among these are our finding(23) that the variations in cosmic-ray intensity are much more closely correlated with the earth's average (or "geomagnetic") latitude than with the geographic latitude. Similarly, Clay has found that the decrease in intensity at the equator is more prominent at longitudes about 120° E. than at longitudes 0 to 90° W., in accordance with the fact that the earth's magnetic field is stronger in the eastern hemisphere(43). These details can leave no doubt but that the latitude effect is due to the action of the earth's magnetic field.

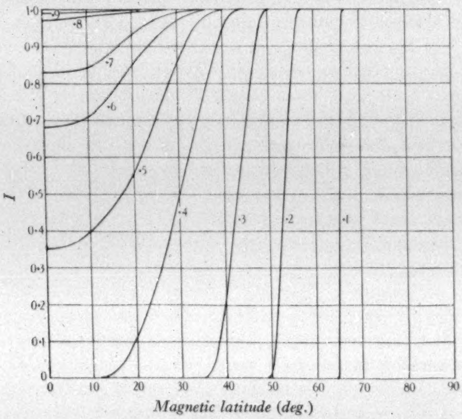


Figure 4. Intensity as a function of latitude calculated for particles of different energy, by Lemaitre and Vallarta.

In its bearing on the composition of cosmic rays, a most significant aspect of the latitude effect is its rapid increase with increasing altitude. This became evident from the high-mountain measurements at different latitudes on our cosmic-ray survey(25), and has been confirmed and extended to higher altitudes by the airplane measurements of Bowen, Millikan and Neher(26) and Clay(27), and especially by a comparison of the stratosphere balloon observations of Regener(28), Piccard and Cosyns(29), Clay(27), and Compton, Stephenson and Millikan(30). In figure 6 are shown some of these data, chosen because of the similarity of the methods of measurement. They show the striking fact that whereas at sea-level the latitude effect is relatively small, near the top of the atmosphere the ratio of the intensity of the rays near the equator to that near the poles seems to be of the order of only 1 per cent.

Clay notes that the values got on his balloon flights in Java are not of high precision. By comparison with his more reliable airplane observations, however, it would seem improbable that his balloon data are in error by more than 20 or 30 per cent. From the shape of his curve it appears that but little increase in intensity is to be expected at the equator above 15 km., and thus that at the top of the atmosphere the intensity of the cosmic rays should be not far from 12 standard ions. The measurements made with the unshielded chamber on our Century of Progress balloon flight, on the other hand (curve 52°, figure 6), showed an intensity of 500 standard ions at a barometric pressure of 5 cm. and increasing with altitude in such a way that a linear extrapolation to the surface of the atmosphere, figure 6, would give about 800 ions. Analysis of our {altitude, ionization} data shows a marked effect ascribable to the earth's field for altitudes over a barometer of 35 cm., so that near the magnetic pole the intensity should be considerably greater, probably more than 2000 ions (cf. figure 7). Thus at the highest altitudes at which observations have been made (at a pressure of about 5 cm. of mercury) the measured ratio of the cosmic-ray-intensity at 52° to that at 0° is about 500/12 or 42. At the surface of the

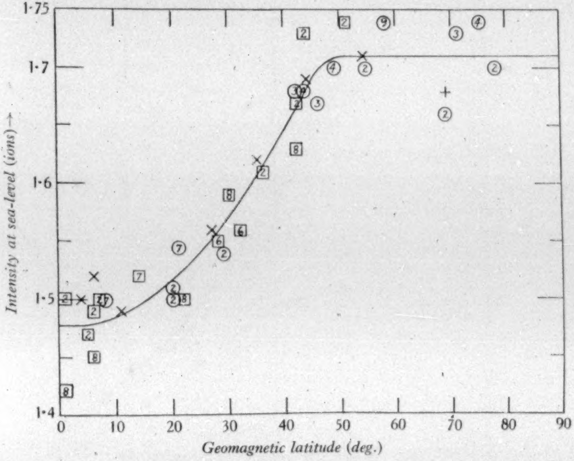


Figure 5. Cosmic-ray ionization at sea-level as a function of geomagnetic latitude. The numbers refer to our different expeditions.

atmosphere the ratio of intensities between the poles and the equator is certainly greater, and probably more than 100 : 1.

Since electrically neutral rays should be unaffected by the earth's field, this result means at once that probably not more than 1 per cent of the ionization at the top of the atmosphere near the poles is due to electrically neutral rays. This possible fraction of 1 per cent is reduced to a still smaller value if we note that the rays received at the equator show a marked east-west asymmetry, indicating that the earth's magnetic field affects a large portion even of those rays which reach the earth at the equator. Thus only a very small fraction indeed of the primary cosmic rays can be electrically neutral.

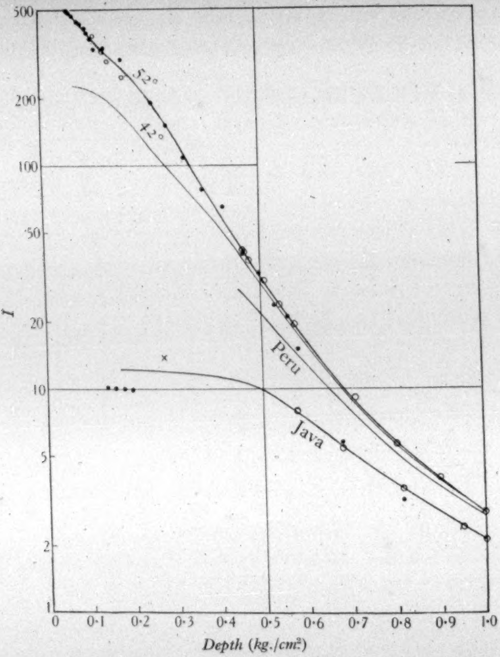


Figure 6. Cosmic-ray ionization as a function of depth below the surface of the atmosphere at different geomagnetic latitudes. 52°, Compton-Millikan; 42°, Millikan-Bowen; Peru, Millikan-Neher; Java, Clay.

The coincidence experiments described above, however, show a shower-producing component of cosmic rays that seems identifiable as photons. This finding can be reconciled with our conclusion of electrically charged primaries if we assume that the photons are secondary rays, excited when primary particles strike the atmosphere. From the fact that the shower-producing rays are relatively less abundant at lower depths we may infer that the primary rays which excite them are less penetrating than the coincidence-producing electrical particles that constitute the chief component at sea-level. There thus appears to be an electrically charged component of the primary cosmic rays which excites the shower-producing photon as it traverses the atmosphere and is more absorbable than the total cosmic-ray beam.

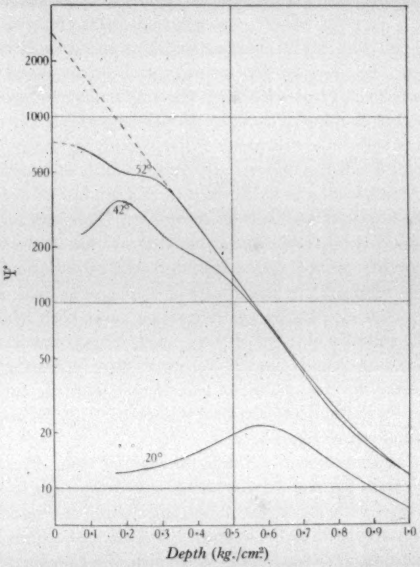


Figure 7. Intensity at different depths for radiation coming vertically through the atmosphere, as observed at different latitudes. 52°, Compton-Stephenson-Millikan; 42°, Bowen-Millikan; 20°, Clay.

The large east-west asymmetry found by Rossi(8) for the coincidence-producing rays that penetrate 8 cm. of lead is further evidence of the electrical character of the more penetrating rays. There thus appears no reason to doubt Bothe and Kolhörster's conclusion that such coincidences are due to the direct action of a component of the primary rays.

Other lines of evidence, based on the form of the {depth, ionization} curve, the independence of the transition effect of latitude, etc., have also led(31) to the conclusion that electrically neutral particles are negligible in the primary cosmic rays. In view, however, of the apparently conclusive character of the evidence from the latitude effect at high altitudes, it seems unnecessary to elaborate the argument.

The need for further consideration of photons, neutrons or other neutral particles as an important part of the primary cosmic rays seems thus to be eliminated(32).

§ 4. MASS ANALYSIS OF THE PRIMARY COSMIC RAYS

Our problem is now reduced to that of identifying the various electrically charged components which may be present in the cosmic rays incident upon the earth. For particles of lower energy such an analysis may be performed by using a mass-spectrograph, in which electric and magnetic fields are employed. Attempts to deflect cosmic rays with laboratory electric and magnetic fields have recently met with some success. The energies of the primary particles are, however, so high that these particles are deflected only with the greatest difficulty; and even when they are deflected, it is hard to distinguish between the primary cosmic-ray particles and the secondaries excited within the atmosphere. Fortunately, however, Nature has supplied us with a ready-made magnetic spectrograph suitable for analysing the primary cosmic rays. The earth itself acts as the magnet, and in place of the electric field we have the stopping-power of the earth's atmosphere. This natural instrument has the advantage of such great dimensions that the rays are analysed far above the atmosphere, where they cannot become confused with secondaries. It leaves something to be desired regarding the uniformity of its magnetic field, and we have not as yet been able to learn accurately the calibration curve with which to determine the energies of the particles in terms of their penetrating-power in the atmosphere. In spite of these limitations and even with the incomplete information now available, an attempt to analyse the components of cosmic rays with our earth magnet leads to valuable results(33), and indicates the kind of data that must be obtained if such an analysis is to be made more rigorous.

Calculation of minimum ranges for different types of rays. From the Lemaitre-Vallarta curves given in figure 4 we may choose a limiting value of x_0 such that at a given latitude particles with greater x_0 will be transmitted, whereas those with smaller x_0 will not be transmitted. Actually, as figure 4 indicates, this limit is not sharp. In table 1 I have accordingly chosen a mean x_0 between the Lemaitre-Vallarta limit of complete transmission and the Störmer limit of complete obstruction. Though no precise calculation of this intermediate region has been made, the values here given cannot differ greatly from the limiting value of x_0 for the rays passing vertically through the atmosphere.

Table 1. Minimum energies at different latitudes

Magnetic latitude (degrees)		0	20	30	40	45	50	55	60	65
Limiting $\bar{x}_0$		0·55	0·49	0·41	0·33	0·28	0·23	0·18	0·14	0·10
Minimum energy (e.V. × 10^9)	± Electrons	18·1	14·3	10·0	6·5	4·6	3·1	1·95	1·15	0·60
	Protons	17·1	13·4	9·1	5·6	3·8	2·32	1·21	0·56	0·17
	Alpha particles	32·5	25·0	16·8	9·7	6·3	3·5	1·73	0·71	0·18
	Oxygen nuclei	130·0	100·0	67·0	39·0	25·5	14·0	6·9	2·2	0·7

The quantity x_0 is defined by

x_0

$$x_0 \equiv \pm r \frac{mv}{eM} \qquad \text{......(1)},$$

where r is the radius of the earth, m the relativity mass of the particle, v its velocity, e its charge, and M the magnetic moment of the earth. From this expression, Lemaitre and Vallarta have tabulated values of the kinetic energy of electrons, protons and alpha particles for various values of x_0. In table 1 are given the corresponding minimum values of the kinetic energies of these particles, and also of oxygen nuclei, for various latitudes. In comparing these figures with experiment, the greatest error is probably that due to the irregularities in the earth's magnetic field, which make uncertain the effective latitude and value of x_0 at a given point on the earth's surface.

r, m, v
e, M

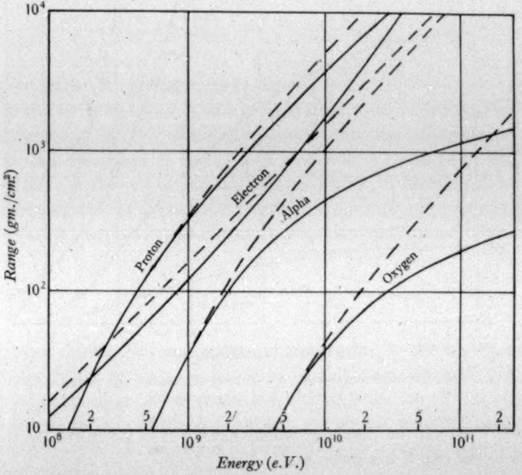


Figure 8. Estimates of the ranges for different energies, for four types of particles. In the preferred estimates (solid lines) account is taken of radiation and collisions with electrons and nuclei.

Corresponding to each value of the minimum energy of a particle there will be a minimum value of its effective range. Unfortunately our information is scanty regarding the relationship between range and energy for particles with cosmic ray energies. Figure 8 gives however in the solid and broken lines two different estimates of this relationship for each of our four particles. We shall consider later (see § 5) how these estimates are made. For the moment we note merely that the estimate represented by the solid line is to be preferred, and that the difference between the two estimates is an indication of the large uncertainty in our knowledge of the ranges of particles with such great energies.

Using the curves of figure 8, we find a minimum range corresponding to each of the minimum energies of table 1. Thus we obtain figure 9, which gives the minimum effective ranges corresponding to different latitudes for each type of particle. The significance of this figure is that, to about the degree of precision of the range curves of figure 8, at any chosen latitude the particles entering the earth's atmosphere may have ranges greater but not less than those listed in figure 9.

Experimental range-minima. A comparison of these predicted minimum ranges with the results of experiment is complicated by the fact that the cosmic rays enter the atmosphere from all directions above the horizon. Gross(34) has however shown
I, z that if I is the observed intensity at a depth z below the surface of the atmosphere,

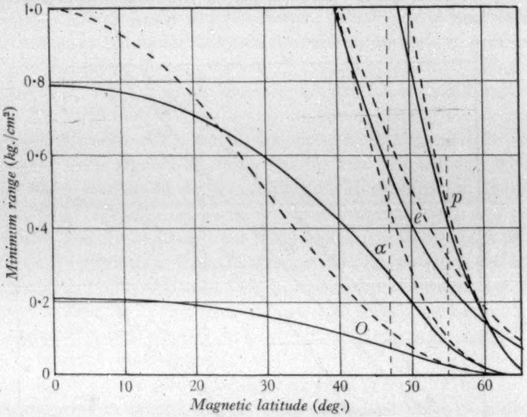


Figure 9. Estimated minimum ranges for different latitudes, for protons, electrons, alpha particles, and oxygen nuclei.

Ψ the intensity Ψ which the radiation would have if it came normally through the atmosphere to the depth z is given by

$$\Psi = I - z\,(dI/dz) \qquad ......(2).$$

By use of this transformation equation, the data collected in figure 6 give the curves shown in figure 7, which represents therefore the intensity of the radiation passing vertically through the atmosphere.

In order to interpret these curves in terms of ranges, consider the following idealized cases:

(i) Range particles, which ionize uniformly along their path until they reach the end of their range, and then stop. High-speed beta and alpha particles approximate to this type, though both show somewhat stronger ionization near the end of their range.

(ii) End-ionizing particles, which have a definite range but produce most of their ionization near the end of their path. In the limiting case they would give no ionization except at the end of the range. A particle which produces most of its ionization through the secondaries it excites will, if it has a definite range, approximate to this type.

(iii) Particles which are absorbed exponentially, and produce ions where they are absorbed. This means that there is a constant probability per cm. of path that the particle will be stopped, but that there is no energy spent by the particle except in the region of absorption. High-speed electrons losing energy in nuclear collisions at which most of their energy is radiated, and alpha particles suffering destructive nuclear encounters, as well as photons photoelectrically absorbed, would fall approximately into this category.

If the absorption is of the exponential type, Eckart(35) has shown how curves like those of figure 7 may be analysed into absorption spectra by use of Laguerre functions. Over the altitude range above 7 kg./cm.², where the curvatures of the $\{\log \Psi, z\}$ curves are chiefly downward, such an analysis must however lead to impossible negative intensities(36). Such an exponential analysis would thus lose its physical significance if applied to these curves.

If, as in the limiting case (ii), all the ionization were to occur at the end of the range, the curves of figure 7 would give directly the range-distribution of the particles, since the ionization at each depth would be a measure of the number of particles arriving at that point. In this case, however, we should expect the ionization Ψ to approach zero at the top of the atmosphere, which it apparently does not do.

If, as in case (i), the ionization is uniform throughout the range, it can be shown(37) that the intensity I_z of the particles having ranges between z and $z+dz$ is given by

$$I_z\,dz = -z\frac{d\Psi}{dz}\,dz \qquad \ldots\ldots(3).$$

I_z

In figure 10 are shown the range spectra corresponding to the curves of figure 7, as calculated according to this assumption. In this case, since all the ranges are presumably greater than some minimum, the slope of the $\{\Psi, z\}$ curves should approach zero at the top of the atmosphere, and should nowhere become positive. Though the former condition seems to be satisfied, the uppermost portions of the Ψ curves show positive slopes.

It is not surprising that none of our idealized cases is capable of representing accurately the cosmic-ray ionization. This ionization is undoubtedly the result of a complex system of primary and secondary rays, each of which ionizes in several different ways. It is apparent from the above discussion, however, that in the upper reaches of the atmosphere the ionization is less like that appropriate to exponential absorption than like that due to particles with a definite range. We may in any case use equation (3) and figure 10 for giving the *effective* range-distribution as if by uniformly ionizing particles. These effective ionization ranges may however be slightly greater than the effective ranges of the primary particles, as used in figures 8 and 9, in view of the additional range of the secondary radiation which may contribute to the ionization.

In figure 11 are collected the Ψ curves calculated from all of the available balloon data for which observations to a sufficient altitude have been made. On the basis of the data from which his curve in this figure was calculated, Kolhörster first called attention(38) to the break in the intensity values at about 0·5 kg./cm^2, which corresponds to our peak *B* of figure 10. For the data from the unshielded meter used on our Century of Progress flight, two Ψ curves are given. The solid line is identical with that in figure 7, being based on the smooth curve drawn through the datum points of figure 6. The dotted line is the result of a similar Gross transformation applied to the smoother curve drawn through the same points by Bowen, Millikan and Neher(39), who did not recognize any irregularities. I have preferred the solid curve, because it represents the datum points more accurately. As will be seen from the corresponding dotted curve of figure 10, however, the range analysis of Bowen, Millikan and Neher's curve leads to limiting ranges of about the same values as given by our curve, though less sharply defined. Similarly the close-dotted curve representing Bowen and Millikan's combined airplane and balloon data near 42° is taken from their smooth $\{I, z\}$ curve(40) in which no irregularities were recognized. The corresponding range curve plotted in figure 10 shows prominently both peaks *A* and *B*. The curve ascribed to Regener-Gross is identical with that of figure 7, and is merely a copy of Gross's own analysis(34) of Regener's average high-altitude data. Piccard and Cosyns(29) called attention to the departure of their data from an exponential curve in the neighbourhood of our *B* peak. The curve representing their data here was calculated(31) from a smooth curve drawn through all their datum

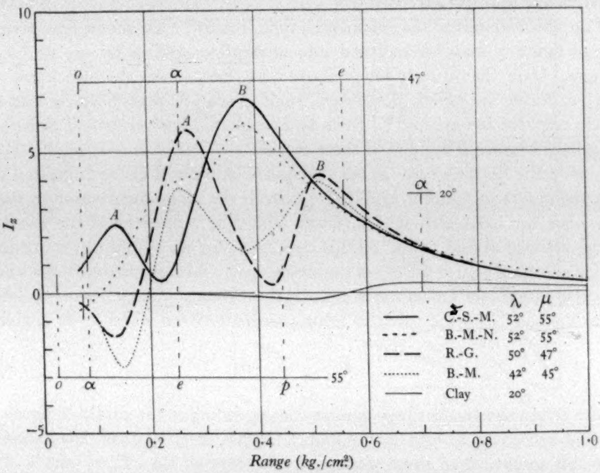


Figure 10. Effective ionization-range distribution of the cosmic-ray particles traversing the atmosphere.

points. That for Clay's data is taken from the corresponding smooth curve of figure 6. The curve for Compton and Stephenson's data from the Century of Progress balloon flight, being taken with a meter shielded by 6 cm. of lead(31), is not strictly comparable with the others. It is of interest however in showing the *B* peak prominently at about the same depth as that found with Millikan and Neher's unshielded chamber. There would thus seem to be no reason to doubt the validity of the breaks in the experimental curves which correspond to the range-distribution peaks *A* and *B*.

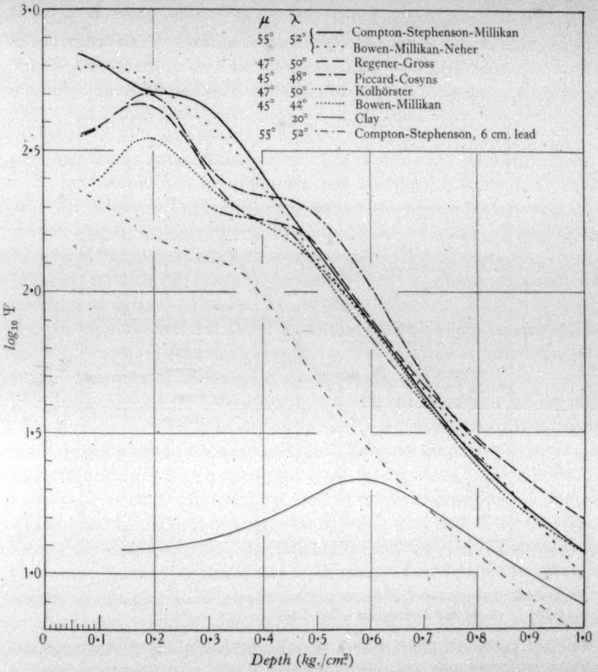


Figure 11. Comparison of vertical intensities as found on various balloon flights, showing systematic variations which depend upon latitude.

For the most part the breaks in the curves of figure 11 show a regular progression in depth with change of the geomagnetic latitude λ. An exception is found, however, in the Bowen-Millikan curve for $\lambda = 42°$, in which the breaks occur at about the same depth as for the three curves obtained in Europe at about $\lambda = 49°$. That this exception is due to the non-uniformity of the earth's magnetic field becomes clear from a consideration of the two curves of figure 12, which show the latitude effect as observed respectively in North America and Europe(41). In both cases the intensity is plotted against the geomagnetic latitude(42). It will be seen however that the knee of the curve at which the latitude effect begins is at about 51° in North America, but at 57° in Europe. This is related to the longitude effect observed by Clay(43), and is doubtless connected with the fact that the local magnetic pole is located in North America(42).

For our purpose it is most convenient to take account of this difference by using a corrected magnetic latitude in terms of which the knee of the latitude-effect curves will come at the same place in both hemispheres. This may be done by assigning the mean value of 54° to the knee of both curves, and correcting the other geomagnetic latitudes by adding three degrees to those in America and subtracting three degrees from those in Europe. In this way we obtain the corrected magnetic latitudes shown in figure 11 under the heading μ. In terms of this corrected value, it will be seen that the humps in figure 11 show a regular progression with latitude. We shall use these corrected magnetic latitudes μ rather than the geomagnetic latitudes λ, since they are thus found from figure 12 to give better correlation with the cosmic-ray intensities.

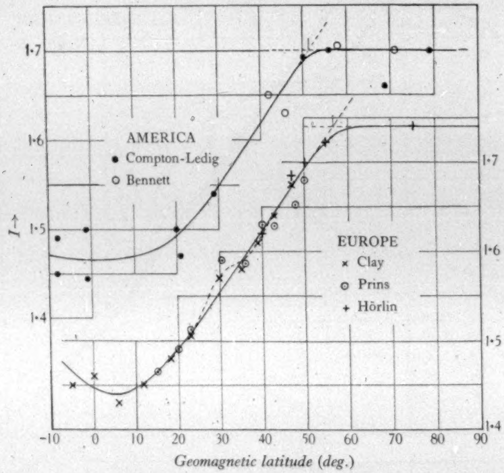


Figure 12. Cosmic-ray-intensity at sea-level as a function of geomagnetic latitude in North America (above) and Europe (below).

μ, λ

Identification of the range-groups. In order to identify the types of particles responsible for an observed range-group, we read from figure 9 the minimum ranges to be expected at the appropriate latitude, and see which of these ranges, if any, fits the minimum range of the observed group, as marked by the sloping part of the curve in figure 10. Thus we note from figure 9 that the preferred estimate of the minimum range of an electron at 55° is 0·25 kg./cm.^{2} This lies on the upward sloping portion of peak *B* of the 55° curve in figure 10. Since the minimum ranges for the other particles do not come in the same region, this would identify peak *B* as due to a group of incoming electrons. In a similar way peak *A* may be identified with alpha particles. For the 47° curve as indicated in the figure, the same assignment of types of particles works best; but the anticipated minimum ranges do not fit as accurately on the upward slopes of the range-distribution curves. The same result is obtained on testing every curve of figure 11: peak *A* corresponds best with alpha-particle ranges, and peak *B* with electron ranges.

A similar test can be made directly from the sea-level {intensity, latitude} curves shown in figure 12. We have seen that the latitude effect begins at about 54°. From figure 9 we find that at sea-level (1 kg./cm.^{2}) the latitude effect should begin at about 48° for protons and at 40° for electrons, and should not occur at all for alpha particles. Considering the uncertainties involved, the agreement is reasonably good with protons, but is unsatisfactory for any other kind of particle. Also, if peak *B* of the 55° curve of figure 10 is ascribed to electrons, we cannot explain the beginning of the sea-level latitude effect at this latitude by the same kind of particle. In the European data of figure 12, which are the more complete, there is an indication of a second break in the curve at about 30° latitude, which might be ascribable to the electrons(43·1). An interpolation between the curves of figure 10 shows that at about this latitude of 30° the *B* peak should appear at about sea-level.

If the 54° knee of the latitude-effect curve is due to protons, it is clear that they should not give rise to a break in any of the {altitude, ionization} curves, since none of these have data taken at a higher latitude. There are however two other possible origins of a latitude effect at a depth too great for the slowest electrons and alpha particles to penetrate: (i) Since a large part of electron absorption is due to excitation of radiation in photons of energy comparable with that of the electron, there must occur extensive straggling, approaching the condition of exponential absorption. Thus some electrons with the minimum energy should penetrate much farther than the average range. (ii) The particles responsible for range-groups *A* or *B* may excite a secondary radiation more penetrating than the primary particles, which would reflect at sea-level the latitude effect on the primary particles at high altitudes. Both of these interpretations should however result in a more gradual onset of the latitude effect than the experiments seem to show at about 54°. It would thus seem that the sharp beginning of the latitude effect at a latitude too high to be ascribed to either group *A* or *B* must mean that it is due to a third range-group of rays, which we may call *C*.

On the basis of our preferred estimates of the ranges of particles of high energies, it is thus possible to identify these range-groups *A*, *B* and *C* respectively as alpha particles, electrons and protons.

§ 5. SOURCES OF ERROR

As will have been noted, there are three relationships which must be known in order to carry through this analysis of the composition of the primary cosmic rays. These are: (i) the minimum energy of a particle penetrating the earth's magnetic field as a function of the geographic position; (ii) the range of the particle as a function of its energy; and (iii) the ionization at representative geographic positions as a function of depth below the surface of the atmosphere.

From the discussion of figure 11, it would appear that incomplete as the cosmic-ray data may be, definite range limits are identifiable, and the values of the limiting ranges calculated from independent {altitude, ionization} measurements differ by much less than do the estimated ranges for different types of particles. High-altitude measurements, especially at different latitudes and approximately the same longitudes, would make our knowledge more complete. Yet errors in the existing ionization measurements can hardly be large enough to change the result of the analysis of the composition of the cosmic rays.

In calculating the minimum energy of various types of particles capable of traversing the earth's magnetic field, however, considerable errors may occur. In the first place, the theoretical calculations of the orbits for the latitudes under consideration are only approximate. For our calculation we have used the value of x_0 at which Lemaitre and Vallarta's curves give 50-per-cent transmission. Though their curves are only approximate, it seems unlikely that a more precise calculation could change the result by an amount comparable with the factor of more than 2 by which the predicted ranges of the various types of particles differ from each other. Much more serious is the assumption that the earth acts as a uniformly magnetized sphere. The fact that the local magnetic latitude(42), which is calculated from the dip of the magnetic needle on the assumption of a uniformly magnetized sphere, is on the average nearly the same as the geomagnetic latitude, supports the approximate correctness of the assumption of uniform magnetization. Attention has been called above, however, to notable departures from an exact correlation between cosmic-ray-intensity and geomagnetic latitude. Had the above calculations been based upon the geomagnetic latitude, the differences between the results from the data of Regener and those of Bowen and Millikan would have been so large as to make the identification of the range-groups ambiguous. Our method of correcting the latitude on the basis of sea-level cosmic-ray measurements is supported by the fact that Fritz's map(44) of the frequency of the occurrence of aurora likewise places Pasadena and Stuttgart at nearly the same effective latitude. At best, however, our method of finding the effective latitude is only approximate. A more detailed study of the theory, taking into account the irregularities in the earth's field, is needed before errors from this source can be considered negligible.

The most important possibility of error however is in the relation between the energy and the range for the various types of particles. Unfortunately we do not possess direct range-measurements for any kind of particle with an energy of the magnitude considered in this discussion. Until such measurements are made our conclusions must be considered tentative. There are nevertheless several guides for making estimates of these ranges.

Up to several million electron-volts, the ranges of electrons, protons and alpha particles have been precisely measured, and formulae for their absorption have been developed⁽⁴⁵⁾. These formulae can be applied with confidence also to oxygen nuclei, and may be used as a basis for extrapolation to somewhat higher energies. Up to about 3×10^8 electron-volts, Anderson has secured data on the absorption of electrons in lead. These data have confirmed a theory of Bethe and Heitler⁽⁴⁶⁾ up to about 7×10^7 e.V., where the theory ceases to be valid, and indicate the direction in which the theoretical predictions are to be modified for still higher energies.

In the broken lines of figure 8, the ranges have been calculated on the following bases: For electrons, an estimate by Bethe of 250 ions per cm. in air has been used as a mean ionization by very high energy electrons, which is equivalent to $6{\cdot}2 \times 10^6$ e.V. gm.$^{-1}$ cm.^{2} This estimate includes the effect of radiation as well as collisions with electrons. For protons, where radiation should theoretically be negligible, and the speeds are low enough for relativity electrodynamics to be applied with confidence, the formulae applicable to particles with lower energies are used. These take into account only collisions with the electrons in the matter traversed. The calculations for alpha particles and oxygen nuclei are made on the same basis.

In forming the estimates shown in the solid lines of figure 8, the calculations have been refined as follows: For electrons, it is noted, in accord with Bethe and Heitler's theory as modified by Anderson's experiments⁽⁴⁶⁾, that the rate of energy-loss rises to a maximum of perhaps 300 ions per cm. in air, falling to about 200 at the higher energies, as provisionally estimated by Bethe. For the heavier particles an attempt has been made to take nuclear collisions into account. Assuming the effective collision area⁽⁴⁷⁾ of an alpha particle to be about 3×10^{-25} cm.^{2}, there should appear an exponential type of absorption which should become of importance at about 100 gm. cm.$^{-2}$ range. Such nuclear collisions have been assumed to become less important at very high energies, to be more prominent for oxygen nuclei, and much less prominent for protons⁽⁴⁷⁾. In every case for low energies the ranges are made to approach those known from experiments in radioactivity.

It is difficult to form an estimate of the probable error of these calculated ranges. Certainly the estimates are less reliable for the higher energies than for the lower ones; but fortunately, except for protons, the {altitude, ionization} data give information regarding ranges for the lower energies of figure 8. In this region it appears improbable that the estimated ranges should be in error by enough to confuse the different types of particles. We should hardly ascribe to chance, therefore, the results of our analysis. Rather, in spite of the very considerable uncertainties involved, we may ascribe to the greatly different limiting ranges anticipated for particles of different types the apparent success of the method.

Conversely, if we are able to identify the particles associated with the various range-groups at high altitudes, experiments of the type considered here will give us more exact information regarding the ranges or absorption of the particles for greater energies. Thus on the basis of the present experiments, we should conclude that electrons of energy greater than 10^9 e.V. have somewhat shorter ranges and protons somewhat longer ranges than are given by the solid curves of figure 8.

§ 6. COMPARISON WITH OTHER DATA

The directional experiments of Johnson (48), Alvarez (49), Rossi (50) and others have shown more rays coming from the west than from the east. This is attributed to the curvature of the paths of positively charged particles in the earth's magnetic field. Johnson and Street's observation (48) that even at a geomagnetic latitude of 50° there is a slight westward preponderance would seem to mean that range-group *C* is positively charged, since this is the only group of primary rays which should show the asymmetry at sea-level at this latitude. This is in agreement with the above conclusion that this group consists of protons.

At the equator, both groups *B* and *C* as observed at sea-level must be strongly affected by the earth's magnetic field. Though the directional asymmetry here is very marked, this of itself does not necessarily mean that both components are positively charged. In fact both Rossi (51) and Clay (24) infer from analysis of their directional experiments near the equator that though positive particles predominate, some negatives are also present. Johnson's earlier analysis (48) seemed to leave little room for negative particles; but his more recent studies of the directional effect (52) have led him also to infer the existence of some negative particles in the primary cosmic rays. Either result would be consistent with the present analysis. We should definitely anticipate a predominance of positive particles because of the proton component *C*; but the electron component *B* might consist of positrons, negatrons, or very possibly about equal parts of each.

Confirming the result that the more penetrating particles observed at sea-level consist of protons is Rossi's observation (53) that the coincidence-producing particles that penetrate 8 cm. of lead show much greater directional asymmetry than do the unfiltered coincidence particles (54). Whether the electron component reaching the earth is positive or negative or both, since it must have greater energy to traverse the atmosphere than would the proton component, it should be less affected by the earth's field. According to Bethe and Heitler's theory (46), however, the difference in the absorption of electrons and protons, being due chiefly to excitation of radiation, should be even greater in lead than in air. Thus the particles transmitted by the lead filter should be predominantly protons, and should accordingly exhibit a large asymmetry, as Rossi's experiments show.

The major characteristics of the shower-producing radiation can on the other hand be explained if we suppose that it consists chiefly of photons excited by the electrons composing range-group *B* (54). We have seen that this group is more absorbable than the proton component *C*. Thus if near the earth's surface the cosmic ray effects are due to the combined direct ionization by *B* and *C* particles plus the ionization by secondary rays from the *B* particles, at higher altitudes this ionization by the secondary rays should be relatively more prominent. This would account for

Rossi's and Johnson's observations(8, 9) that at higher altitudes the shower-producing radiation is relatively more abundant, and also the fact that at higher altitudes the transition effect is more prominent. Moreover, recent studies by Johnson(52) show that at high altitudes near the equator, where range-group *B* should be relatively prominent, the shower-producing rays have very little directional asymmetry. When compared with the fact that these rays show a marked latitude effect, this demands for their origin primary rays which consist of about equal parts of positive and negative particles. This fits exactly with our conclusion that the shower-producing rays are connected with the electron component *B*, but adds the information that these electrons are about equally positive and negative.

Regener and Pfotzer(55) have recently shown that the curve of impulses obtained with a tube counter at various altitudes up to 28 km. is the same in shape as that obtained with the ionization chamber. Their conclusion is that the specific ionizing power of the cosmic rays is practically the same for the whole region investigated. At first thought this might be taken as evidence that at all altitudes the composition of the cosmic rays is the same. These measurements were however made with a single unshielded tube which would respond to any ionizing particle capable of penetrating the walls. Experiments with Wilson chambers, counter tubes and ionization chambers all indicate that at sea-level practically the whole ionization is due to high-speed particles of about the specific ionizing power of high-speed electrons, most of which particles are secondaries. This should be true whether the primary cosmic rays are negatrons, positrons, protons or photons. If they are alpha particles, some increase in the specific ionization might have been expected; but a low specific ionization is not difficult to explain. The most obvious suggestion is that as soon as a nuclear collision, which might well be the most important type of absorption, occurs the alpha particle may disintegrate into component protons, electrons, and perhaps neutrons. Since the cosmic-ray energies are greatly in excess of the packing-effect energy-decrement of the alpha particle, about 3×10^7 e.V., it would in fact be remarkable if the alpha particle could survive such a collision. After such disintegration, the greater part of the original energy would be spent in ionization by the component parts of the nucleus, which would show about the normal specific ionization appropriate to an electron.

In their Wilson-chamber experiments, Blackett, Anderson and others have hardly been able to measure energies appropriate to the primary cosmic-ray particles, though some of these must show themselves in their photographs. Anderson concludes(56) that the large majority of his tracks are secondaries, about equally divided between positives and negatives, but that of the particles having more than 10^9 volts energy (which would include most of the primaries) there appears to be some excess of positive particles. This result seems rather to confirm the deductions drawn above. On the other hand it is rather surprising that these experiments have never shown definite evidence of protons, which might be distinguished by their relatively high specific ionization if moving slowly enough, whereas our analysis assigns protons to the range-group *C* which is important at sea-level. It may be that all of the protons which are able to traverse the earth's magnetic field have too great a range to be stopped by the atmosphere, and traverse the Wilson chambers at a speed so high that their specific ionization is about the same as for an electron. Or is it possible merely that the chances are against catching a photograph of a proton trail near enough to the end of its range to make its greater specific ionization evident? In any case we must at present consider this apparent absence of proton trails on the Wilson photographs as a notable difficulty in the way of accepting the results of our analysis. If on the other hand cosmic alpha particles are to be detected by the Wilson method, the experiment should be performed high in the stratosphere, and no existing photographs offer evidence as to their existence.

We should note also the surprising observation, made by Lenz(57) by means of his new electrical method of deflecting cosmic-ray particles, that most of the very high-energy particles are negative. This seems to be in definite disagreement with the results of the Wilson photographs in strong magnetic fields which, as we have seen, seem to indicate a preponderance of positives. In both cases the particles concerned are probably mostly secondaries, whereas the directional experiments, which indicate a definite excess of positive particles, refer to primaries before they enter the atmosphere. As yet Lenz's results are of a preliminary character, and we shall wait with interest to see whether his further experiments continue to show the excess of high-energy negatrons.

§ 7. RESULTS OF THE ANALYSIS

Practically all of the primary cosmic radiation consists of electrically charged particles. This is shown by the fact that measurements of the cosmic rays at very high altitudes near the equator show only 1/40 as great an intensity as is observed at corresponding altitudes at about 54° magnetic latitude. Apparently this ratio would be less than 1/100 at the top of the atmosphere if the intensity at the magnetic equator and poles were to be compared. There is thus room for less than 1 per cent of photons or other electrically neutral particles in the primary cosmic rays.

Counter-tube experiments, as shown by Rossi and others, reveal the presence of two distinct types of cosmic radiation which reach the earth at sea-level. One consists of highly penetrating electrical particles, and causes coincident impulses between counters placed in line. The other produces showers of absorbable particles, and apparently consists of photons. Both components show a marked latitude effect. The photonic shower-producing radiation is thus considered to be a secondary radiation excited by charged particles (probably electrons) as they traverse the atmosphere. The penetrating coincidence particles, on the other hand, appear to be primary protons and electrons.

The method of analysing the cosmic rays here described indicates the presence of three groups of particles, characterized by having at a given latitude three distinct minimum ranges. The three range-groups are designated A, B and C in the order of their ranges. Best agreement between the minimum energy admitted through the earth's magnetic field and the minimum observed ranges is found if group A is identified with alpha particles, group B with electrons (either positrons or negatrons as far as this analysis goes), and group C with protons. When the evidence from the directional experiments is considered, it appears that group B is probably composed about equally of negatrons and positrons. There is no evidence for particles heavier than helium atoms.

Further {altitude, ionization} measurements should refine the argument, but are not expected to alter the identification of the range-groups. It is not impossible, however, that uncertainties in the effective value of the earth's magnetic field and especially in the relation between the energy and range of different types of particles may be sufficient to lead to false identification of the range-groups. The generally satisfactory agreement for observations at different geographic locations, however, lends some confidence to the results.

This analysis of the composition of the primary cosmic rays seems to agree well with our information regarding the directional asymmetry of the rays, and with the variation of the shower-producing component with altitude. There is some difficulty, however, in accounting for the failure of the Wilson photographs of cosmic rays to show protons, which our analysis determines as an important component of the rays at sea-level.

REFERENCES AND NOTES

(1) W. Bothe and W. Kolhörster, *Naturwissenschaften*, **16**, 1045 (1928); **17**, 271 (1929), *Z. Phys.* **56**, 751 (1929).

(2) R. A. Millikan, *Phys. Rev.* **43**, 661 (1933).

(3) J. C. Street and T. H. Johnson, *Phys. Rev.* **42**, 142 (1932).

(4) C. D. Anderson, *Phys. Rev.* **44**, 406 (1933).

(5) B. Rossi, *Z. Phys.* **68**, 64 (1931); B. Rossi and G. Bottecchia, *Ricerca Scient.* **5** (I), 171 (1934). See also Rossi in the *Proceedings of the London Conference on Nuclear Physics* (1934), where he discusses these and other similar experiments leading to conclusions consistent with those here drawn.

(6) D. S. Hsiung, *Phys. Rev.* **46**, 653 (1934).

(7) B. Rossi, *Proc. Lond. Conf. on Nuclear Phys.* (1934).

(8) Reference (7) and B. Rossi and S. de Benedetti, *Ricerca Scient.* **5** (I), 594 (1934) and **5** (II) (Dec. 1934).

(9) T. H. Johnson, *Phys. Rev.* **45**, 569 (1934). Note, however, C. W. Gilbert, *Proc. roy. Soc.* **144**, 559 (1934), who finds no such difference, though new measurements by Johnson (52) seem to establish the results described in the text.

(10) J. C. Street and T. H. Johnson, *Phys. Rev.* **42**, 142 (1932).

(11) Cf. Rossi (7) for a summary of this evidence, with references. Also C. D. Anderson, R. A. Millikan, S. Neddermeyer and W. Pickering, *Phys. Rev.* **45**, 352 (1934).

(12) J. Clay, *Proc. Acad. Sci. Amst.* **30**, 1115 (1927); **31**, 1091 (1928); also **33**, 711 (1930).

(13) R. A. Millikan and G. H. Cameron, *Phys. Rev.* **31**, 163 (1928).

(14) W. Bothe and W. Kolhörster, *Berlin Ber.* No. 26, p. 450 (1930).

(15) Kerr Grant, *Nature*, Lond., **127**, 924 (1931).

(16) R. A. Millikan, *Phys. Rev.* **36**, 1595 (1930).

(17) Cf. e.g. G. Hoffmann, *Phys. Z.* **32**, 633 (1932).

(18) E.g. R. A. Millikan (2) who ascribes Clay's differences to a temperature effect.

(19) C. Störmer, *Z. Astrophys.* **1**, 237 (1930); *Oslo Obs. Publ.* No. 10 (1934).

(20) P. S. Epstein, *Proc. nat. Acad. Sci.*, Wash., **16**, 658 (1930).

(21) B. Rossi, *Nuovo Cim.* **8**, 3 (1931). E. Fermi and B. Rossi, *Acc. Nat. dei Lincei*, **17**, 346 (1933).

(22) G. Lemaitre and M. S. Vallarta, *Phys. Rev.* **43**, 87 (1933).

(23) A. H. Compton, *Phys. Rev.* **43**, 387 (1933), gives a summary account of the findings up to 1933. A. H. Compton and P. G. Ledig, *Phys. Rev.* **45**, 294 (1934). For later data, cf. A. H. Compton, *Trans. Amer. geophys. Un.* p. 154 (1933).

(24) J. Clay and H. P. Berlage, *Naturwissenschaften*, **20**, 687 (1932). H. Hoerlin, *Nature*, Lond., **132**, 61 (1933); *Naturwissenschaften*, **21**, 822 (1933). J. Clay, *Physica*, 's Grav., **1**, 363 and 829 (1934). I. S. Bowen, R. A. Millikan and H. V. Neher, *Phys. Rev.* **44**, 246 (1933); **46**, 641 (1934). J. Prins, *Nature*, Lond., **132**, 781 (1933). P. Auger and M. L. Rinquet, *Proc. Lond. Conf. on Nuclear Phys.* (1934). B. Rossi, I. Ranzi, S. de Benedetti, *Ricerca Scient.* **5** (1), 575 (1934).

(25) A. H. Compton, *Phys. Rev.* **41**, 111 (1932); **43**, 387 (1933).

(26) I. S. Bowen, R. A. Millikan and H. V. Neher, *Phys. Rev.* **44**, 248 (1933).

(27) J. Clay, *Physica*, 's Grav., **1**, 363 (1934).

(28) E. Regener, *Phys. Z.* **34**, 306 (1933), and elsewhere.

(29) A. Piccard and M. Cosyns, *C. R. Acad. Sci.*, Paris, **195**, 605 (1932). M. Cosyns, *Nature*, Lond., **135**, 313 (1935). Cosyns notes a marked effect, of the order of 5 per cent per degree, as his balloon drifts from one latitude to another.

(30) Dr Stephenson and I wish to thank Prof. Millikan and Dr Neher for accepting our invitation to supply one of the two cosmic-ray meters used on the Century of Progress flight, and for calculating and making available to us the results of the ionization measurements made with it. In their publication of these data (*Phys. Rev.* **46**, 641 (1934)) Bowen, Millikan and Neher do not mention that we had any share in obtaining them. It was primarily for securing the stratosphere measurements with a shielded and an unshielded ionization chamber that Dr F. R. Moulton and I initiated the Century of Progress balloon flight, and it was at the expense of slighting the rest of the elaborate scientific programme that Commander Settle, Major Fordney and Dr Stephenson were able to make these two experiments successful (cf. A. H. Compton, *Proc. nat. Acad. Sci.*, Wash., **20**, 79 (1934)). Perhaps its initiation, planning, direction and performance, and months of undivided effort toward its succesful accomplishment will justify including the names of Dr Stephenson and the writer in describing this experiment.

(31) A. H. Compton and R. J. Stephenson, *Phys. Rev.* **45**, 441 (1934).

(32) Though the conclusion that primary photons play a negligible role in cosmic-ray phenomena undoubtedly reflects the opinion of the large majority of investigators in this field, two groups of investigators have recently expressed views differing from the conclusions here reached: (i) H. Kulenkampff (*Phys. Z.* **30**, 561 (1929)) and Regener, Kramer and Lenz (*Z. Phys.* **85**, 411 and 435 (1933)), have interpreted the flattening of the upper part of the {depth, ionization} curve (figure 6) as due to photons entering the atmosphere, which produce ionization only after they have transferred a part of their energy to recoil electrons. The absence of this flattening in the high-altitude measurements taken nearer the magnetic pole shows, however, that the effect at lower latitudes is due primarily to the action of the earth's magnetic field on the incoming electrified particles. (ii) Bowen, Millikan and Neher(30) conclude "that nearly all of the non-field-sensitive part of the ionization of the atmosphere above sea-level is due to photons of energy 200 ± 170 million electron volts". They estimate that "93 per cent of the sea-level ionization is due to non-field-sensitive rays", which seems to imply that much the greater part of the sea-level ionization is due to *primary* photons. In reaching this conclusion these authors have neglected entirely the evidence from coincidence experiments, and have overlooked the great 100 : 1 latitude effect at high altitudes. A detailed discussion of their arguments would be out of place here.

(33) An analysis of the type here presented was made by Compton and Stephenson(31) on the basis of less complete data, leading to similar but less definite results.

(34) B. Gross, *Z. Phys.* **83**, 217 (1933).

(35) C. Eckart, *Phys. Rev.* **45**, 851 (1934).

(36) If the intensity is the sum of two exponential components,

$$\Psi = \Psi_1 e^{-\mu_1 z} + \Psi_2 e^{-\mu_2 z}$$

we have for the curvature of the $\{\log \Psi, z\}$ curve:

$$\frac{d^2}{dz^2} \log \Psi = {}^{1}\frac{\Psi\Psi_2}{\Psi^2}\, e^{-(\mu_1+\mu_1)z}\, (\mu_1 - \mu_2)^2.$$

This is positive for all real values of μ_1 and μ_2 unless either Ψ_1 or Ψ_2 is negative.

(37) Gross(34) shows that the number of "range particles" entering the atmosphere, as measured in terms of the ionization at the surface, is

$$R_z\, dz = -(d\Psi/dz)\, dz.$$

Since the total ionization along the path is the ionization per cm. × the range, we have for the total ionization by the particles of ranges between z and $z + dz$,

$$I_z dz = zR_z dz = -z\,(d\Psi/dz)\, dz \qquad \text{......(3).}$$

If the observed ionization I is expressed in ions per cm.³ per sec. in standard air (as in figure 5), $I_z dz$ is the total number of ions produced per cm.² per sec. by cosmic rays having ranges between z and $z + dz$. It is thus proportional to the intensity, or energy per cm.² per sec., of the rays of these ranges.

(38) W. Kolhörster, *Verh. dtsch. Phys. Ges.* **16**, 719 (1914).

(39) Bowen, Millikan and Neher's curve(30) given in their figure 7 was used for this Gross transformation.

(40) Their curves(30) in figures 1 and 6 were used.

(41) For American data, cf. *Trans. Amer. geophys. Un.* p. 154 (1933). For European data, cf. reference (24). The data have been multiplied by suitable factors (nearly unity) to make them coincide at $\lambda = 40°$.

(42) We follow writers on terrestrial magnetism who distinguish between the "geomagnetic latitude", referred to the pole of the earth's uniform magnetization, and the "magnetic latitude", defined by

$$\tan \Lambda = \tfrac{1}{2} \tan \delta \qquad \text{......(4),}$$

where δ is the inclination of the magnetic needle. The pole of the earth's uniform magnetization is located at 78° 32′ N., 69° 08′ W.; whereas the north magnetic pole ($\delta = 90°$) is at 70° 30′ N., 95° 30′ W. (cf. *Smithsonian Physical Tables*, 8th ed., p. 575).

(43) J. Clay, *Physica*, 's Grav., **1**, 363 and 829 (1934). Clay notes that this effect "may be completely explained by the fact that the axis of the earth's magnet is situated at a distance of about 300 kilometres from the centre of the earth". The same phenomenon has also been very recently reported by R. A. Millikan and H. V. Neher, *Phys. Rev.* **47**, 205 (1935).

(43.1) Similar breaks are present near the same latitude in the just-published data of Millikan and Neher (*Phys. Rev.* **47**, 205 (1935).

(44) Cf. C. Störmer, *Les Aurores Boreales* (Paris, 1925), plate 25, credited to Fritz.

(45) Cf. Rutherford, Chadwick and Ellis, *Radiations from Radioactive Substances* (1930), pp. 101 et seq. and 414 et seq.

(46) H. Bethe and W. Heitler, *Proc. roy. Soc.* A, **146**, 83 (1934). New data by C. D. Anderson, *Proc. Lond. Conf. on Nuclear Phys.* (1934) have appeared which give Bethe and Heitler's theory a more adequate test than was possible to them. In figure 13 Anderson's measurements of the energy loss per cm. in lead, made on nine tracks before and after traversing lead plates, are plotted as the solid circles. The open circle represents the experimental range of a 3,000,000-volt β particle. The broken line represents Bethe and Heitler's theoretical values, account being taken only of electron collisions. The solid line is for their theory, account being taken also of the energy spent in exciting radiation (their table II). Bethe and Heitler give $137mc^2 = 7 \times 10^7$ e.V. as about the maximum energy at which their theory should be valid. It will be seen that up to this energy the agreement between the experimental

data and the theory is reasonably satisfactory. The somewhat lower values of the datum points is, as Bethe and Heitler point out, due in part to the method of calculating the experimental values, which is valid only for energy-losses small compared with the initial value. The large variations in the experimental values also support the theory in indicating that if any energy is lost by radiation, it is usually a considerable fraction of the energy of the β particle.

For energies greater than 7×10^7 e.V., the data seem to indicate a marked reduction in the rate of energy-loss. On the other hand, the Wilson photographs show ionization at approximately the same rate along the path of the electrons, even for the highest energies that have been measured, and Bethe and Heitler's theory indicates that for these high energies the radiation processes should be of relatively great importance. On the other hand, for electrons traversing air the radiation should be relatively much less prominent than for those traversing lead.

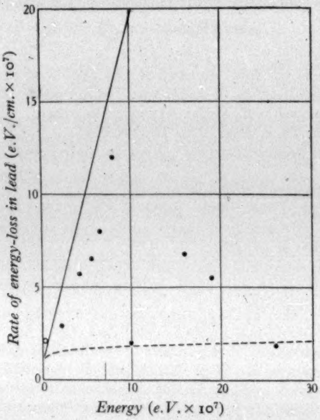


Figure 13. Energy loss per cm. by high-energy electrons traversing lead. Experiments Anderson, theory Bethe and Heitler.

We note also Kolhörster and Tuwim's estimate of 135 ions per cm. in air as the average energy loss by the cosmic-ray particles traversing the atmosphere (*Z. Phys.* **81**, 435 (1933)). This is based on the ratio of the total ionization observed in an ionization chamber to the number of particles traversing it as counted with coincidence tubes. This figure is a mean for all penetrating particles (apparently both protons and electrons) which reach the earth, and must be smaller than that for primary electrons considered alone.

(47) Rutherford, Chadwick and Ellis(45) (p. 255) give $3{\cdot}5 \times 10^{-13}$ cm. as the effective combined radii of an alpha particle and a proton colliding with energies of the order of 10^6 volts. If the nuclei of the atoms in air formed impervious barriers of this radius against the high-energy cosmic-ray particles, their absorption coefficient in air should be about 16×10^{-3} gm.$^{-1}$ cm.^{2} Actually the penetrating component in Eckart's analysis (*Phys. Rev.* **45**, 851 (1934)) has an absorption coefficient of about $0{\cdot}6 \times 10^{-3}$ gm.$^{-1}$ cm.^{2}, i.e. about 4 per cent of the calculated value. This must mean that if nuclei are of the supposed magnitude, they must be readily penetrable by cosmic-ray particles. For protons moving with nearly the speed of light there is no reason to suppose that the nuclei should be much less permeable than for electrons. For the heavier and slower

alpha particles and oxygen nuclei, however, we may expect such collisions to play a more important role. Unfortunately existing data give little basis for calculating its importance.

(48) T. H. Johnson and J. C. Street, *Phys. Rev.* **43**, 381, and T. H. Johnson, *Phys. Rev.* **43**, 834 (1933); **44**, 856 (1933).

(49) L. Alvarez and A. H. Compton, *Phys. Rev.* **43**, 835 (1933).

(50) B. Rossi, *R. C. Accad. Lincei*, **13**, 47 (1931); *Ricerca Scient.* **4** (I), 365 (1933); **5** (II), 579 (1934).

(51) B. Rossi and S. de Benedetti, *Phys. Rev.* **45**, 214 (1934).

(52) T. H. Johnson, *Phys. Rev.* (in press), MS. kindly shown to the writer. His earlier experiments(9) had shown a greater asymmetry for the shower-producing rays at sea-level.

(53) B. Rossi and S. de Benedetti, *Ricerca Scient.* **5** (I), 594 (1934).

(54) The interpretation here given of Rossi's result and of the origin of the shower-producing radiation have been proposed and further elaborated by A. H. Compton and H. A. Bethe, *Nature*, Lond., **134**, 734 (1934).

(55) E. Regener and G. Pfotzer, *Nature*, Lond., **134**, 325 (1934).

(56) C. D. Anderson, *Proc. Lond. Conf. on Nuclear Phys.* (1934).

(57) E. Lenz, *Nature*, Lond., **134**, 809 (1934).

REVIEWS OF BOOKS

An Introduction to Atomic Physics, by J. Thomson. Pp. x + 228. (London: Methuen and Co., Ltd.) 10*s*. 6*d*. net.

This book gives a short, logically arranged outline of atomic physics. It is divided into three parts of almost exactly equal lengths, of which the first summarizes the experimental basis of the subject. Part II, on the theory of atomic structure, naturally opens with the conventional elementary account of Bohr's treatment of hydrogenic atoms and with a short chapter on atoms with many electrons. This is followed by an excellent summary of Hamiltonian methods, leading to an account, illustrated by simple applications, of the wave mechanics of Schrödinger and de Broglie. Part III contains further applications, mainly qualitative, to "Molecular, Atomic and Nuclear Radiations".

The number of minor errors and mis-statements is rather large. Some clearly result from attempts at a too rigorous compression or simplification of the argument, but some must be ascribed to simple carelessness. For instance, we find (p. 3) the not uncommon confusion of a gramme with a gramme-equivalent of hydrogen. On p. 18 an intruding "not" completely stultifies a sentence. In Bucherer's experiments (p. 19) the apparatus was *not* "placed between the large poles of an electromagnet"; Bucherer used a large solenoid—at least, he says he did, and he publishes a photograph of the solenoid at the end of his paper. Further, it is not true that in these experiments each β particle after leaving the condenser describes the arc of a circle. The separation of the isotopes of neon was first demonstrated by the parabola method, not by the mass spectrograph (p. 28). The atomic *numbers* of the isotopes of copper are *not* 63 and 65 (p. 33). The photoelectric work function (p. 41) is not "characteristic of the substances composing the cathode *and anode*", though the measured stopping potential-difference is. There are other errors of a similar kind.

Many of these errors are trivial, but some might well worry the very type of reader for whom in other respects the book is so admirably suited—namely, the "serious student" of physics, that legendary figure who plays a part in the vocabulary of book-reviewing which is closely analogous to that of the magnetic shell in electromagnetism and of perfectly reversible processes in thermodynamics. On the other hand, the book has many features which are worthy of the highest commendation; it is very well planned as an introduction to its subject, and it is written with an unusually clear appreciation of the points which require special emphasis and in a manner which is calculated to stimulate thought on the part of the reader.

H. R. R.

Données Numériques de Spectroscopie, Spectres d'Émission, by L. Bruninghaus; *Spectres d'Absorption*, by V. Henri; *Électro-magnéto-optique*, by F. Wolfers; *Diffusion de la Lumière*, by P. Auger. Pp. xxii + 284. (Extracted from vol. x (1930) of *Tables Annuelles de Constantes et Données Numériques*. Paris: Gauthier-Villars, 1932.)

As most users of the *Annual Tables* are now aware, after the appearance of each volume special sections of it dealing with certain subjects, such as spectroscopy, are separately bound and issued as extract volumes, so that specialists who do not have access to the complete volumes may obtain collected data in their own subjects. The plan is, of course, a great boon and, in the case of spectroscopy at least, would be a still greater if the long period between the appearances of the complete volume and the corresponding extract volume were considerably reduced. The last volume of spectroscopic data, which appeared in 1932, contained some 1400 pages extracted from volumes 8 and 9 (1927–9) of the complete work, and we have had to wait two years for the present one which, being extracted from volume x (1930) only, is very much smaller than its predecessor. By far the largest of the four sections of the book is that on emission spectra which occupies 204 pages, the next on absorption spectra having only 42 pages. From the reader's point of view perhaps a better arrangement would be a division of these two sections into three dealing with both emission and absorption spectra of (i) atoms, (ii) diatomic molecules and (iii) tri- and poly-atomic molecules. With such a rearrangement we should not have, for example, the undesirable separation of twenty pages between the O_2 emission data (pp. 530–2) and the O_2 absorption data (pp. 552–3); again, the emission data for benzene and its derivatives (p. 514) would adjoin the fluorescence data (p. 539) instead of being almost lost amongst fifty pages of data for diatomic molecules (pp. 487–536). A curious double error occurs in the note on BeH (p. 554): the paper cited was by Watson, not Bergstein, and the band was observed in emission, not absorption. Fortunately the variety of notations used in the papers from which the notes and tables of this volume are taken is much less bewildering than it was in the case of the previous extract volume, since the now universally adopted notations in some branches of the subject were just coming into use in 1930. This collection of data, like its predecessors, will be, of course, a very valuable adjunct to the recent books on atomic and molecular spectra. We look forward to the appearance, without undue delay, of the spectroscopic extracts from volumes XI and XII.

W. J.

Physical and Dynamical Meteorology. By DAVID BRUNT. Pp. xxii + 411. (Cambridge University Press.) 25*s*. net.

A few months ago David Brunt became Professor of Meteorology in the University of London in succession to Sir Gilbert Walker. The excellence of the appointment will be appreciated by the readers of the treatise on *Physical and Dynamical Meteorology* which has recently been published. The aim of this book is to give the student of meteorology a thorough knowledge of the physics of the atmosphere. The practical application of this knowledge to weather problems is not considered until quite a late stage in the work. The text covers 403 pages and the first weather map is on p. 315. Before that page is reached the author has dealt fully with the thermodynamics of the atmosphere, with radiation, with the general equations of motion, with turbulence and with transformations of energy. It is to be remarked however that he has forgotten to deal with the nature of condensation. The words "nucleus", "raindrop" and "ice crystal" do not occur in the index of subjects. There are other striking omissions, e.g. thunderstorms and fog.

Perhaps the chapter of most general interest is the long one of 53 pages on "The polar front and its relation to the development of cyclones." This chapter is open to one criticism. When the polar-front theory was first developed the Norwegian school thought of the front as a circle surrounding the pole with east winds on the north of the circle and west winds to the south. A cyclone was generated as a disturbance of this system of winds. Only the most careful of readers would gather from Brunt's account of the matter that the first diagram which he gives is merely of historical interest. The diagram shows a front with cold and warm air flowing along it in opposite directions. There is an approximation to such a state of things in the case of a trough of low pressure, but it is not in such troughs that cyclones generally originate. The more orthodox treatment of the subject with a depression formed on the boundary between two currents in the same direction but of different temperatures follows after a few pages, but the inexperienced student is not warned as he should be that the earlier diagram is not to be taken seriously.*

---

* The reader should be warned that in the most elaborate example of formal analysis in the book, figures 82 and 83 have been interchanged by the printer.

The last chapter of the book is devoted to "The general circulations of the atmosphere". It is strange that whereas the theory of the trade winds used to be regarded as a matter to be disposed of in a few lines in an elementary text-book of geography, Brunt has to sum up his discussion by saying that "it is not possible at present to put forward any satisfactory theory of the general circulation". In fact the mathematician is not yet in a position to say "this is why the trade winds blow within the tropics, this is why we have prevailing west winds in middle latitudes and that is the mathematical analysis which justifies my statements and is fully supported by observation".

There are other big gaps in theoretical meteorology. The most conspicuous is the lack of an explanation of the fact that the lowest temperatures in the atmosphere occur in the stratosphere over the equator. The mathematical physicist who takes up the study of meteorology under the guidance of Prof. Brunt will realize how much solid work has been done, but he will realize also how much scope there is both for more flashes of genius and for more patient investigation.

F. J. W. W.

General Astronomy. Second Edition, by H. SPENCER JONES. Pp. viii + 437. (Arnold.) 12*s*. 6*d*. net.

The appearance of a new edition of the Astronomer Royal's *General Astronomy* is very welcome. The book has been reset throughout, and so great has been the progress in the last 12 years that it is practically a new work. Everywhere we find evidence that the latest papers have been examined and incorporated. The author recalls that in the first edition (1922) the discussion of theories of the spiral nebulae was summed up in the words "Although the balance of evidence at present seems opposed to the island-universe theory, the question cannot be regarded as yet definitely closed". To-day this reads like an echo of the middle ages.

To cover satisfactorily the whole of modern astronomy in 430 pages can only be accomplished by a miracle of compression. The allotment of space is: atronomical instruments, methods, etc., 100 pages; solar system, 170 pages; stars and nebulae, 150 pages. In no other work is so much information concerning both the facts and theories of astronomy set forth in so concise a way. But, although encyclopaedic in character, the book is always readable. There is no elaborate attempt at simplified exposition, but the layman should have no difficulty in understanding.

It is as a source of reference that the book is most valuable. One hesitates to use the well-worn word "indispensable". But when the latest information is needed on some question of astronomy which stands rather aside from the more popular researches, this is the natural work to consult; and it is rarely that we are disappointed.

A. S. E.

(1) *A five year bibliography of the theory of refrigeration, refrigerants and appliances.* (2) *A five year bibliography of the applications and testing of refrigeration and of its British patents.* The Science Museum, South Kensington. Compiled by H. T. PLEDGE. Published by H.M. Stationery Office. Price: printed on one side of the page the cost of each bibliography is 2*s*. 6*d*., printed on both sides 2*s*.

When the Science Museum organized the special exhibition dealing with refrigeration in 1934 steps were taken to compile bibliographies of the subject extending over the period 1929–1933. These volumes should prove of permanent value to those interested in refrigeration and the Museum authorities are to be congratulated on the enterprise.

E. G.

THE EFFECT OF PRESSURE ON SUPERSONIC DISPERSION IN GASES

By W. RAILSTON and E. G. RICHARDSON

Page 533 *of this volume*

DISCUSSION

Prof. E. N. da C. Andrade. I am very interested to see that the authors find that the velocity of sound for CO_2 has a minimum in the supersonic region, falling as the frequency increases beyond a certain value. This is not in accord with Kneser's theory, but agrees with the results obtained with air in my laboratory by Mr E. B. Pearson.* As I pointed out in connexion with Mr Pearson's paper,† the regional absorption for CO_2 is not consistent with Kneser's theory and would lead one to expect the kind of variation of velocity which the present authors have shown to exist. This experimentally established variation, exhibited by both air and CO_2, points to some mechanism of resonant absorption, and this, although it has not yet been theoretically explained, may exist in addition to the mechanism contemplated by Kneser, which is undoubtedly operative as well.

Author's reply. We are pleased to hear that Prof. Andrade favours our suggestion that some form of resonant absorption takes place in carbon dioxide at 100 kc./sec. and that the work of Mr Pearson in his laboratory suggests a similar mechanism in air. As a matter of fact, Kneser in his own experiments obtained but few results above this frequency, whereas those workers whose results extend up to 1000 kc./sec. have mostly found a fall of velocity in carbon dioxide and nitrous oxide at the highest frequencies, although they have not specially remarked upon the fact. Certainly the experimental results are not so simple in explanation as the relaxation time theory would have us believe.

---

* *Proc. phys. Soc.* **47**, 136 (1935).

† *Loc. cit.* p. 147.

THE PROCEEDINGS OF
THE PHYSICAL SOCIETY

VOL. 47, PART 5 *September* 1, 1935 NO. 262

534·323

A CORRECTION TO THE THEORY OF THE RAYLEIGH DISC AS APPLIED TO THE MEASUREMENT OF SOUND-INTENSITY IN WATER*

BY A. B. WOOD, D.SC., F.INST.P.

ERRATA

"A negative-resistance device and its application to harmonic analysis", by C. W. OATLEY, M.A., M.Sc., vol. **47**, p. 471.

for 621.3.018.362 1.396.645.22 *read* 621.3.018.3 621.396.645.22

"The effect of pressure on supersonic dispersion in gases", by W. RAILSTON, M.Sc. and E. G. RICHARDSON, B.A., D.Sc., Ph.D. Remarks in discussion by Prof. E. N. DA C. ANDRADE, F.R.S., vol. **47**, p. 777.

for the velocity of sound for CO_2 has a minimum in the supersonic region
read the velocity of sound for CO_2 has a maximum in the supersonic region

ellipsoid of revolution in a vibrating fluid. In the case of an oblate ellipsoid of revolution (a disc-like body) he found that the turning-moment T is given by T

$$T = \tfrac{4}{3}\rho a^3 v^2 \sin 2\theta \qquad \text{......(1)},$$

where ρ is the density of the fluid medium, a the radius of disc, v the r.m.s. velocity ρ, a, v of the fluid flow and θ the angle between the normal to the disc and the direction of θ the undisturbed fluid flow. If the disc has appreciable thickness a small multiplying factor $(1-0{\cdot}15\, t/a)$, where t is the thickness of the disc, must be applied to the above t expression for T. Usually this factor approximates closely to unity.

The streamlines of flow round a disc inclined to the undisturbed stream are illustrated in figure 1. A and B are points of zero velocity and therefore of maximum pressure. The manner in which the torque arises obviously requires no explanation. It will be seen from equation (1) that the couple is a maximum when $\theta = 45^\circ$ and becomes zero when $\theta = 0^\circ$ or 90°, i.e. when the disc is edge-on or broadside-on to the flow. The couple T tending to bring the disc broadside-on to the flow is proportional to v^2, a quantity independent of the direction of flow. Consequently T may be used

A
D

A
ap
su
th
fl
in
va
ex
cy
sn

el
re

w
of
th
fa
ex

il
p
It
b
fl
to

THE PROCEEDINGS OF
THE PHYSICAL SOCIETY

VOL. 47, PART 5 September 1, 1935 No. 262

534·323

A CORRECTION TO THE THEORY OF THE RAYLEIGH DISC AS APPLIED TO THE MEASUREMENT OF SOUND-INTENSITY IN WATER

BY A. B. WOOD, D.SC., F.INST.P.

Received March 15, 1935. *Read June* 7, 1935

ABSTRACT. König's theory of the Rayleigh disc assumes the disc to be motionless, apart from its slow rotation. This assumption is unjustifiable in a relatively heavy medium such as water. The disc partakes of the motion of the medium to an extent depending on the density of the medium and the mass and dimensions of the disc. Consequently the fluid-velocity v in König's equation for the torque on the disc must be modified by the introduction of a factor β involving the velocity of the disc relative to the medium. The value of β for ellipsoidal discs is calculated from hydrodynamic theory and determined experimentally. A comparison is made of the results for ellipsoidal discs and for flat cylindrical discs. A method of measuring sound-intensities under water by means of small heavy metal discs is described.

§ 1. THEORY

THE theory of the Rayleigh disc was first developed by W. König[1] in a series of papers entitled "Hydrodynamic Acoustic Investigations". He obtained a general expression for the turning-moment experienced by a stationary ellipsoid of revolution in a vibrating fluid. In the case of an oblate ellipsoid of revolution (a disc-like body) he found that the turning-moment T is given by T

$$T=\tfrac{4}{3}\rho a^3 v^2 \sin 2\theta \qquad \text{......(1)},$$

where ρ is the density of the fluid medium, a the radius of disc, v the r.m.s. velocity ρ, a, v
of the fluid flow and θ the angle between the normal to the disc and the direction of θ
the undisturbed fluid flow. If the disc has appreciable thickness a small multiplying
factor $(1-0{\cdot}15\, t/a)$, where t is the thickness of the disc, must be applied to the above t
expression for T. Usually this factor approximates closely to unity.

The streamlines of flow round a disc inclined to the undisturbed stream are illustrated in figure 1. A and B are points of zero velocity and therefore of maximum pressure. The manner in which the torque arises obviously requires no explanation. It will be seen from equation (1) that the couple is a maximum when $\theta=45°$ and becomes zero when $\theta=0°$ or $90°$, i.e. when the disc is edge-on or broadside-on to the flow. The couple T tending to bring the disc broadside-on to the flow is proportional to v^2, a quantity independent of the direction of flow. Consequently T may be used to measure v^2 in alternating flow, as in a sound-wave, whence the sound-intensity is calculated.

This method has been in use as a standard method for measuring sound-intensity in air for many years and is still relied upon for the calibration of sub-standards of various kinds. In applying the method to under-water sounds, however, a serious difficulty arises. König's theory, summarized above, assumes the disc to be motionless, apart from its slow rotation, and this assumption is unjustifiable in a relatively heavy medium like water. Allowance must be made for the fact that the disc partakes of the motion of the water, and the quantity v in König's equation (1) must define the velocity of fluid flow *relative* to the disc.* Another point implied, but not definitely stated, in König's theory is that the disc must be rigid, i.e. it must not be set into any form of transverse or radial resonant vibration in the range of frequencies over which it may be used. Otherwise the theory would break down completely.

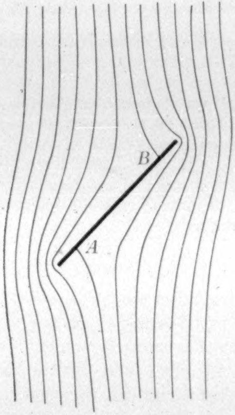


Figure 1. Streamline flow round a thin circular disc inclined at 45°.

v_d, v_w If we represent the r.m.s. velocity of the disc by v_d and of the water by v_w and write

β

$$v_d/v_w=\beta;$$

then the relative velocity v between the water and the disc is given by

$$v=v_w-v_d=v_w\,(1-\beta) \qquad \text{......(2).}$$

---

* It is understood from a private communication that Prof. L. V. King, McGill University, Montreal, working independently on this problem, has arrived at the following conclusions as a result of a comprehensive mathematical treatment. "The velocity potential in the neighbourhood of a Rayleigh disc may be completely expressed in terms of cylindrical wave functions. When the circumference of the disc is small compared to the wave-length, the integral equations involved may be approximately solved by the use of Hankel's inversion theorem. The final formula for the torque on the disc differs from Rayleigh's original formula by an inertia factor due to the oscillations of the disc in the incident wave. The final practical formula applicable to the disc as a sound-measuring device shows the existence of an optimum thickness depending on the radius and relative densities of the disc and of the medium in which the sound-waves are propagated."

Equation (1), which gives the couple tending to turn the disc broadside-on, becomes

$$T=\tfrac{4}{3}\,\rho a^3 v_w^2\,(1-\beta)^2\sin 2\theta \qquad \ldots\ldots(3).$$

We have now to determine the value of β, the ratio of the velocity of the disc to that of the water. If M is the mass of the disc, m that of the water displaced, and m_w the M, m, m_w water load, we have

$$\frac{v_d}{v_w}=\frac{m+m_w}{M+m_w}=\beta$$

and

$$(1-\beta)=\frac{M-m}{M+m_w}.$$

It is shown in text-books of hydrodynamics* that the water load on a thin ellipsoidal disc broadside-on to the fluid flow is $\tfrac{8}{3}a^3\rho_w$. When the disc is inclined at 45° this water load m_w reduces to $\tfrac{4}{3}a^3\rho_w$ approximately, whilst in the edge-on position the water load on a thin disc approaches zero. When the disc is inclined at 45° we have therefore

$$(1-\beta)=\frac{M-m}{M+m_w}=\frac{M-m}{M+\tfrac{4}{3}a^3\rho_w} \qquad \ldots\ldots(4).$$

The intensity (erg/cm.² sec.) in the sound-wave in water is given by

$$I=\rho_w C v_w^2 \qquad \ldots\ldots(5),$$

where C is the velocity of sound in water ($1{\cdot}45\times10^5$ cm./sec. in fresh water at C 15° C.) and v_w is the r.m.s. velocity of the water in the undisturbed sound-wave. v_w Substituting the value of v_w^2 given by equation (3) and assuming $\theta=45°$ we find that this becomes

$$I=\frac{\tfrac{3}{4}CT}{a^3(1-\beta)^2} \qquad \ldots\ldots(6).$$

In this expression for the sound-intensity the torque T on the disc in the 45° position can be measured; the quantities a and β are known from the linear dimensions and mass of the disc, and C is the known velocity of sound in the water. It will be shown later how the value of β can be checked experimentally.

A fine suspension wire attached to a torsion head serves to measure the torque T required to maintain the disc in its initial 45° setting against the couple exerted by the sound wave. If α is the angle of twist, $T=k\alpha$, where k is a constant characteristic α, k of the suspension used to support the disc. Small angular deflections near the 45° position may be observed directly without using the torsion head.

The expression for the sound-intensity given in equation (6) above is of course applicable not only when water is the medium surrounding the disc but also when we are dealing with sounds in air. In the latter case, however, the correction due to the motion of the disc and the mass of air carried with it is small. As an example let us suppose a disc of mica 1 cm. in diameter and 0·002 cm. thick, with M equal to 5×10^{-3} gm., is chosen to measure intensity in air and in water, then the comparative values of the quantities immersed are as shown in table 1.

---

* Lamb, *Hydrodynamics*, pp. 152, 162 (1895).

Table 1

	Air	Water
M	5×10^{-3} gm.	5×10^{-3} gm.
m	2×10^{-6} gm.	$1{\cdot}6 \times 10^{-3}$ gm.
m_w*	2×10^{-4} gm.	0·16 gm.
β	0·038	0·98
$1/(1-\beta)^2$	1·08	2500

* Load due to the medium (air or water).

The quantity $1/(1-\beta)^2$ is the correction factor required to König's formula for the intensity as measured by the Rayleigh disc. It will be seen therefore that whereas a correction of only 8 per cent is required in the indications of a thin mica disc used in air, the same disc in water would require a multiplying factor of 2500. Such a disc is therefore quite useless in water; a much heavier disc is necessary. For example, compare now the corrections required when the disc employed is a tungsten ellipsoid of mass $M=6{\cdot}85$ gm. and diameter 2 cm.; these are given in table 2. In this case the factor $1/(1-\beta)^2$ is quite negligible in air and has been reduced to 40 per cent only in water.

Table 2

	Air	Water
M	6·85 gm.	6·85 gm.
m	$4{\cdot}7 \times 10^{-3}$ gm.	0·36 gm.
m_w	$1{\cdot}1 \times 10^{-3}$ gm.	0·85 gm.
β	9×10^{-4}	0·158
$1/(1-\beta)^2$	1·002	1·4

These two examples serve to show the necessity for using as heavy a disc as possible in water. In air this necessity is far less urgent, but it appears that the correction factor may become important if very light mica discs are used.

The acoustic power-output of any oscillator may be determined when the
I axial intensity I and the beam-shape are known. In certain cases, as for example a
high-frequency circular oscillator of piston type, the beam-shape conforms to a
known theoretical form(2, 3) whence the sound-output is easily calculated in terms
of the measured sound-intensity on the axis. The total acoustic energy-output per
E, S second E of such an oscillator of area S vibrating with r.m.s. velocity v over its
whole surface is given by

$$E=\rho c S v^2 \qquad \text{......(7)},$$

λ provided the diameter of the oscillator is considerably greater than a wave-length λ.
R_c Such a source emits an approximately parallel beam of radiation to a distance R_c,
where

$$R_c=m^2\lambda$$

m, r and $m=r/\lambda$, r being the radius of the circular oscillator. For values of R appreciably
larger than R_c the axial intensity I is given by

$$I=\rho C S v^2 \pi m^2/R^2 \qquad \text{......(8)},$$

whence from (7) and (8)

$$E=\frac{\lambda^2 R^2}{S} I \qquad \text{......(9)}.$$

The value of I is determined by means of the disc at a distance R from the oscillator by means of equation (6) above, whence the power-output E of the sound-source is readily calculated from equation (9).

If the beam-shape of the oscillator does not conform to a standard type it may be necessary to perform a three-dimensional integration to obtain the power-output in terms of the axial intensity.

For a non-directional source whose dimensions are small compared with the wave-length λ, the power-output E is given by

$$E = 4\pi R^2 I \qquad \text{......(10).}$$

§ 2. EXPERIMENTAL MEASUREMENTS

Determination of the water load m_w. The value of the water load on the disc may be simply calculated; in the 45° position of the disc, as we have seen, $m_w = \frac{4}{3}\rho_w a^3$. It is nevertheless of interest and importance to provide an experimental check on the theoretical value. Equation (4) above may be re-written in the form

$$m_w = \frac{M - m}{1 - \beta} - M \qquad \text{......(11).}$$

For sounds of constant intensity the torque on the disc is proportional to $a^3(1-\beta)^2$ as shown by equation (6). For discs of the same radius a, therefore, the angular deflection δ of the disc is proportional to $(1-\beta)^2$ or $\sqrt{\delta}$ is proportional to $(1-\beta)$. Equation (11) may therefore be written

$$m_w = K\left(\frac{M - m}{\sqrt{\delta}}\right) - M \qquad \text{......(12),}$$

where K is a constant.

Consequently if deflections δ are measured for various values of M, the graph of $(M-m)/\sqrt{\delta}$ and M will be a straight line cutting the axis of M at $-m_w$. For discs of the same radius, the value of M may be varied (i) by varying the thickness of discs of the same material, and (ii) by using materials of different densities.

The results of such measurements are shown in figures 2 and 3. In figure 2, cylindrical brass discs of radius 1 cm. and of varying thickness were used. In figure 2*a* the observed values of angular deflection δ are plotted as a function of thickness, which is proportional to M, whilst in figure 2*b* the values of $(M-m)/\sqrt{\delta}$ are plotted against M. A similar pair of curves is shown in figures 3*a* and 3*b*. In this case cylindrical discs of different density but of the same dimensions (radius 1 cm., thickness 0·2 cm.) were used. It will be seen that both figures 2*b* and 3*b* indicate a linear relationship between M and $(M-m)/\sqrt{\delta}$ and yield a value of 1·3 gm. for m_w. Similar observations plotted in figure 5, to which reference will be made later, give a value 1·1 gm. These experimental values agree tolerably well with the theoretical estimate of $\frac{4}{3}$ gm. for ellipsoidal discs of radius 1 cm.

A comparison has also been made of discs of different radii but of the same material and thickness, figure 4. The calculated and observed values are compared in table 3.

Table 3

Diameter $2a$ of disc (cm.)	$1-\beta$	$a^3(1-\beta)^2$	Relative deflections (degrees)	
			Calculated	Observed
1·5	0·748	0·235	1	1
2·0	0·707	0·500	2·13	2·3
2·5	0·675	0·891	3·8	3·8

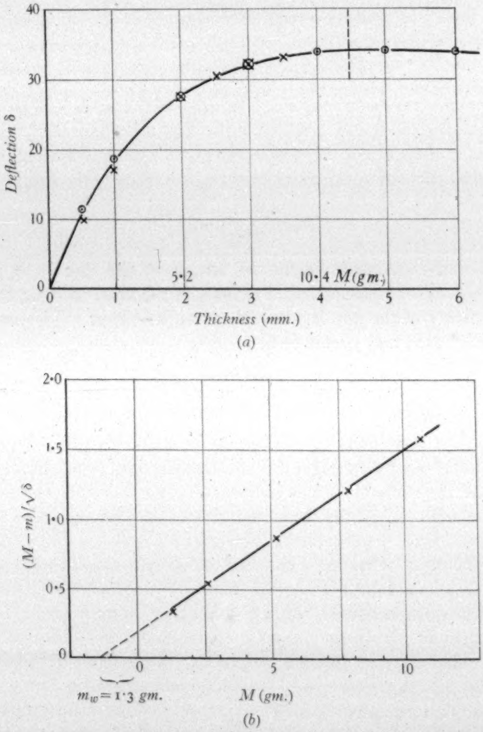


Figure 2. Cylindrical brass discs of radius 1 cm. and of varying thickness.

In the calculation of the fourth column of this table the theoretical value $\frac{4}{3}\rho_w a^3$ has been assumed for the water load on the discs. The agreement between observation and calculation gives support to this value and to the theoretical prediction that the deflection is proportional to the cube of the radius of the disc.

In the above experiments the discs used were all cylindrical and flat. The theoretical deductions of König relate, however, to ellipsoidal discs, and it is of some importance to examine what difference, if any, is introduced by the use of cylindrical

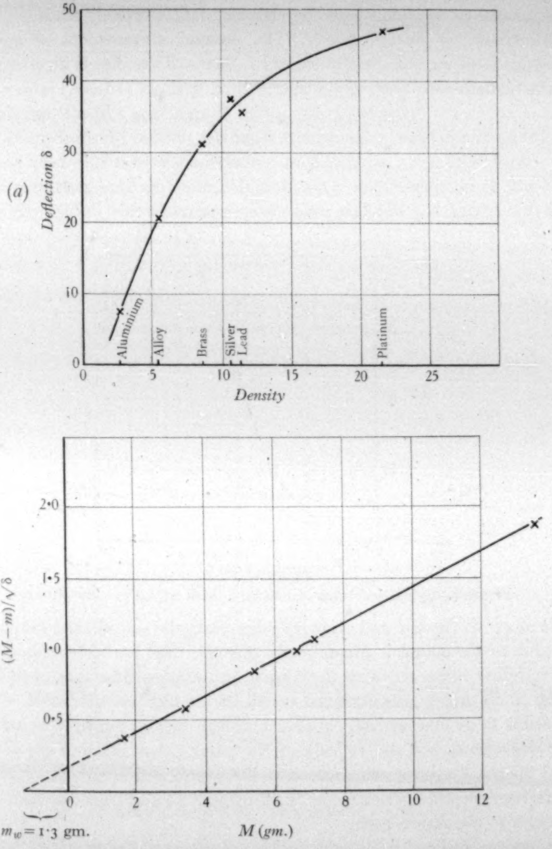


Figure 3. Cylindrical discs of different materials, radius 1 cm., thickness 2 mm.

discs. A comparison was therefore made between a series of cylindrical brass discs of radius 1 cm. and of varying thickness, and a similar series of ellipsoidal discs. The results, plotted in figure 5, indicate that the difference if any is small, the slope of the line and the m_w intercept being practically the same in the two cases. Differences arise when the thickness of the discs exceeds 4 mm. but this would be expected on theoretical grounds.

In accordance with theory, the results shown in figures 2, 3 and 5 clearly indicate the increased sensitivity obtainable by increasing the weight of the disc. The correction factor due to the water load can thus be made relatively small.

Measurement of sound-intensity. The general arrangement of apparatus for measuring sound-intensity under water by means of the disc is illustrated diagrammatically in figure 6. The disc, suspended on a short (10-cm.) wire of phosphor bronze, 0·028 cm. in diameter, is supported from a long stiff rod carrying a torsion head. In the first experiments a spot of light was reflected from the polished surface of the disc, but the surface polish of metal discs is as a rule very poor and this method had to be abandoned. Considerable optical improvement was obtained by using a thin (0·010-in.) stainless steel concave mirror 1 cm. in diameter mounted at an angle of 45° to the disc and arranged edge-on to the sound. The torque on such a mirror due to the sound is theoretically zero and was found by experiment to be quite negligible compared with the torque on the disc. The spot of light reflected from the small mirror was arranged to fall on a white xylonite scale where it was easily visible to an eye vertically above. The scale served not only to indicate small angular deflections, but also to indicate the exact 45° position of the disc. In this position the small mirror was edge-on to the sound and broadside-on to the light, a very convenient arrangement, the beam of light to and from the mirror being at right angles to the sound-path.

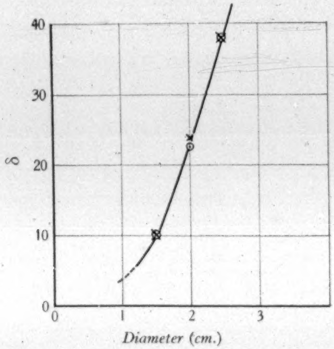


Figure 4. Cylindrical brass discs 2 mm. thick and of varying diameter.

The torsion head controlling the position of disc could be operated by means of a pulley and cord situated conveniently for the observer near the under-water lamp and scale. An inclined mirror over the torsion head was used to reflect towards the observer an image of the pointer and angular scale of the suspension. To reduce the effects of direct flow of water past the disc, due to possible disturbances in the tank, a thin (0·010-in.) clear celluloid cylinder about 6 in. in diameter was arranged as shown in figure 6. This flabby celluloid screen, whilst sufficient to stop slowly-varying water currents, permitted the sound to pass freely, observations with and without the cylinder showing no appreciable difference. Provision was also made for transport of the disc without risk of breaking the suspension—a small metal cone at the lower end of a rod served to raise the disc when required and consequently to remove the tension from the fine suspension wire. Experiments showed that increased damping of the suspended system could be obtained, if required, by means of a thin mica vane mounted symmetrically on the disc-suspension rod and arranged either edge-on or broadside-on to the sound. In these positions the torque on the damping vane due to the sound is zero. Generally, however, the damping due to the

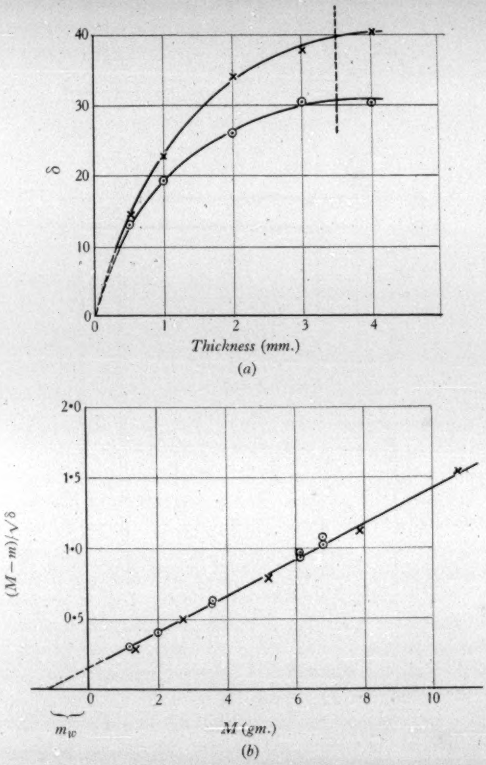


Figure 5. Comparison of cylindrical and ellipsoidal brass discs of radius 1 cm. Ellipsoidal discs ⊙ ⊙; cylindrical discs × ×.

disc and mirror was found sufficient without the addition of the mica damping-vane. The torsion constant of the suspension was determined experimentally by observing the period of torsional oscillation in air when a disc of known mass and radius was suspended from a point on its edge. The torque on the disc is given by

$$T = k\alpha,$$

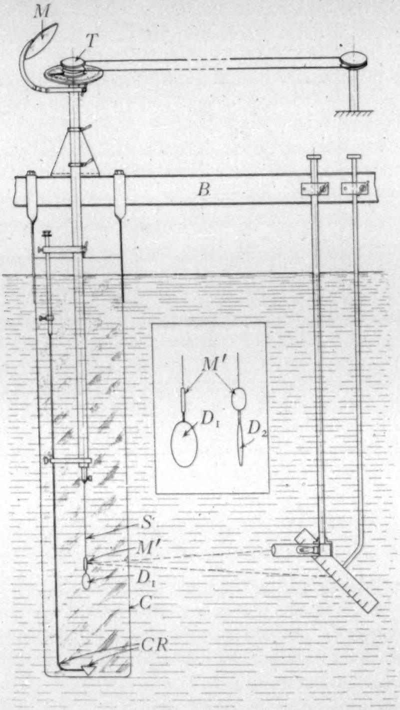


Figure 6. General arrangement. M, M', mirrors; T, torsion head and scale; B, bridge; D_1, disc at 45° to sound; D_2, disc edge-on to sound; CR, clamping rod and cup; C, thin celluloid cylinder; S, suspension wire.

k, τ where k is a constant equal to $\pi^2 Ma^2/\tau^2$, M the mass in grams of the cylindrical disc of radius a, τ the period in seconds of torsional oscillation and α the angle of twist measured in radians.

In using this apparatus to measure sound-intensity in the tank it was soon discovered that small changes of frequency in the source of sound produced appreciable changes in deflection of the disc. This effect was traced to stationary waves formed by reflection from the walls of the tank. The difficulty was surmounted by superposing a small frequency-modulation on the transmitted sound. In this connection it is important also to note that the disc has directional properties, which make it more sensitive to sound arising in some directions than in others. The couple T turning the disc is proportional to $\sin 2\theta$, where θ is the angle between the direction θ of the sound and the normal to the disc. The directional characteristic shown in figure 7 shows maxima at $45^\circ \pm n\pi/2$.

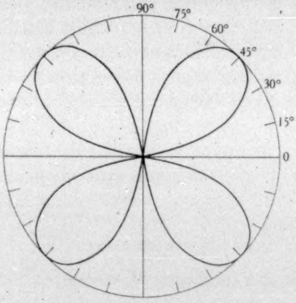


Figure 7. Directional characteristic of disc.

§ 3. SIZE AND FREQUENCY OF TRANSVERSE VIBRATION OF THE DISCS

In the theory of the Rayleigh disc it is stipulated that the circumference of the disc should be considerably less than a wave-length of the sound of which the intensity is to be measured. In water this condition is easily fulfilled since the velocity of sound is high, 1450 m./sec., a sound of frequency 1000 c./sec. having a wave-length of 145 cm. The small discs 1 cm. in radius used in the present experiments are in this respect in accord with the theoretical conditions.

Another factor of importance, which is generally overlooked, relates to the possibility of transverse vibration of the disc in its fundamental mode, with two nodal diameters, or at its first overtone with one nodal circle. When a disc is inclined at 45° to the direction of the sound-wave a pressure-difference exists between the forward and rear edges of the disc, and if the frequency of sound happens to coincide with one of the natural resonant frequencies of the free disc a vibration may be set up which renders the indications valueless. The theoretical frequencies of free circular discs in air are given by Rayleigh* as follows. For a disc vibrating with two nodal diameters (the lowest possible natural frequency of the disc)

$$N_x = 0{\cdot}26\,t\,c/a^2, \qquad N_x$$

where t is thickness and a the radius of the (cylindrical) disc and c the velocity of sound in the material of the disc. For a disc vibrating with one nodal circle

$$N_0 = 0{\cdot}41\,t\,c/a^2. \qquad N_0$$

---

[*] Vol. 1, chap. x. See also Lamb, *Sound*, p. 153.

These expressions refer to flat cylindrical discs. The frequencies of ellipsoidal discs of the same radius and thickness at the centre will of course be considerably higher, owing to removal of mass from the outer edge without much change of stiffness.

In an accompanying paper a description is given of the experimental determination of the frequencies of small circular discs, and it is shown that the experimental values are in reasonably good agreement with the theory. For the present therefore we may use the values of frequency calculated from the above expressions.

The lowest frequency of a cylindrical platinum disc, 2 cm. in diameter and 2 mm. thick, was found to be about 14,000 c./sec., whilst that of an ellipsoidal tungsten disc 2 cm. in diameter and 2 mm. thick at the centre was higher than 20,000 c./sec. The frequencies in water will be slightly lower than these values. Such discs are therefore quite safe to use over a wide range of frequencies extending into the supersonic region.

In this connection it is of interest to calculate the frequency of a small mica disc such as is commonly used to measure sound-intensity in air. If the disc has a radius of 0·4 cm. and a thickness of 0·0025 cm., it could vibrate with two nodal diameters at a frequency N_x where

$$N_x = 1300 \text{ c./sec.}$$

and with one nodal circle at a frequency N_0 where

$$N_0 = 2050 \text{ c./sec.}$$

In the calculation of these frequencies the velocity of sound in sheet mica was taken as $3{\cdot}2 \times 10^5$ cm./sec., a value found experimentally by a method described by the author and Dr F. D. Smith (4). It will be seen that the small mica disc has two natural modes of vibration within the aural frequency range.

Data relating to two discs for use under water. The data incorporated in table 4 relate to a cylindrical platinum disc and an ellipsoidal tungsten disc which have been used in the measurement of sound-intensity under water.

Table 4

	Platinum disc, cylindrical	Tungsten disc, ellipsoidal
Radius	1·00 cm.	0·863 cm.
Thickness (at centre)	0·20 cm.	0·216 cm.
Frequency (2 nodal diameters)	14 kc./sec.	20 kc./sec.
Mass M	13·5 gm.	6·85 gm.
Mass of water displaced m	0·63 gm.	0·36 gm.
Calculated water load m_w	1·33 gm.	0·85 gm.
$1-\beta$	0·87	0·842
$1/(1-\beta)^2$	1·32	1·41
v_w^2	$0{\cdot}99\, k\alpha \times \pi/180$	$1{\cdot}64\, k\alpha \times \pi/180$
I (erg/sec.-cm.^{2})	$1{\cdot}44 \times 10^5 k\alpha \times \pi/180$	$2{\cdot}39 \times 10^5 k\alpha \times \pi/180$
I (erg/sec.-cm.^{2}) when $k=0{\cdot}225$*	565 α†	940 α†

* With an annealed phosphor-bronze wire suspension 28 μ in diameter and 10 cm. long, k is 0·225.

† α is in degrees.

§ 4. MEASUREMENT OF SOUND-INTENSITY IN WATER. SOUND-OUTPUT OF SOURCES

An example of the type of measurement made by means of the discs described above is illustrated in figure 8. In this case the sound-intensity at a distance of 1 m. from the source of sound is plotted as a function of the square of the value of the alternating current used to excite the source into vibration, i.e. as a function of power-input. It will be seen that the sound-output is directly proportional to the

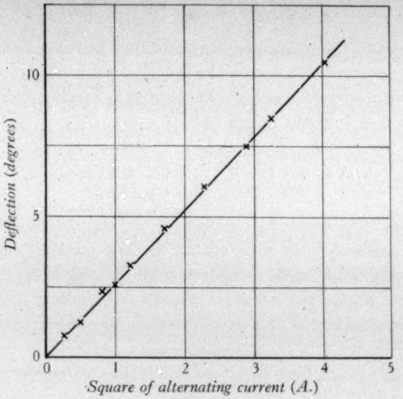


Figure 8.

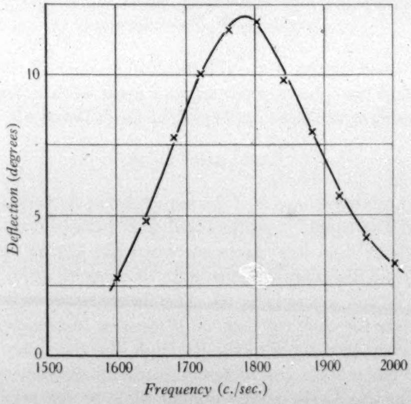


Figure 9.

power-input. With an input current of 2 A., it is found that the sound-intensity at 1 m. is 6800 erg/cm.^{2} sec. Assuming that the sound-intensity is uniform in all directions, then the acoustic power-output of the source is given by

$$\begin{aligned} E &= 4\pi \times 100^2 \times 6800 \text{ erg/cm.}^2 \text{ sec.} \\ &= 85 \times 10^7 \text{ erg/cm.}^2 \text{ sec.} \\ &= 85 \text{ W.} \end{aligned}$$

In another case the disc has been used to determine the frequency characteristic of a sound source under water. This type of measurement is illustrated in figure 9. The intensity of the sound emitted is shown throughout a range of frequencies embracing the resonance frequency of the sound-source.

§ 5. AN ALTERNATIVE TO THE RAYLEIGH DISC—THE DOUBLE SPHERE

W. König$^{(1)}$ not only derived an expression for the torque on a thin ellipsoidal disc, but also for the torque on an elongated prolate ellipsoid, approximating to a cylindrical rod of diameter small compared with its length supported horizontally from the mid point. Such a rod, adjusted initially with its axis at 45° to the direction of the sound-wave, is subjected to a couple tending to rotate it into the broadside-on position.

König$^{(5)}$ also calculated the forces operating on two spheres in an alternating fluid flow and Rayleigh* has pointed out that in the case where the two spheres are rigidly connected together the forces reduce to a couple T where

$$T = 3\pi\rho \frac{a_1^3 a_2^3}{r^3} v^2 \sin 2\theta,$$

θ θ being the angle which the axis of the combination makes with the sound direction,
a_1, a_2, r, v a_1 and a_2 the radii of the spheres, r the distance apart of their centres, v the r.m.s.
ρ velocity of the fluid flow and ρ the density of the fluid. When $a_1 = a_2 = a$ and $r = ma$,

$$T = \frac{3\pi}{m^3} \rho a^3 v^2 \sin 2\theta.$$

It will be seen that this expression is of the same form as that obtained by König for the Rayleigh disc. The couple is proportional to v^2 (which in turn is proportional to the sound-intensity) and is a maximum when the axis of the combination is inclined at 45° to the direction of the sound. The sensitivity of the arrangement varies directly as a^3 and inversely as m^3. The theory requires, however, that the spheres should not be too close together, i.e. m must be considerably greater than 2. In comparing this arrangement with the Rayleigh disc the factor $3\pi/m^3$ in the expression for the torque must be considered in relation to the corresponding quantity $\frac{4}{3}$ for the disc. For the double-sphere combination, m can never be less than 2 when the spheres are in contact and the constant $3\pi/m^3 = 1{\cdot}18$. If $m = 4$, a more likely arrangement, $3\pi/m^3 = 0{\cdot}147$, which is only $\frac{1}{9}$ of the constant for a disc of the same radius.

---

* *Sound*, II, 47.

As a matter of general interest, a few preliminary experiments have been made with two pith balls 0·25 cm. in radius mounted with their centres 1·0 cm. apart. In this case $a=0{\cdot}25$ and $m=4$. The combination was suspended by a thin silk fibre in a tunable resonance tube, the double-sphere arrangement replacing the usual Rayleigh disc. The experimental results were found to agree, within the limits of the rough comparison, with the theoretical value for ratio of sensitivities of the double sphere and the disc.

Although the double-sphere arrangement is always less sensitive and perhaps less convenient in air than the disc, it is just possible that it may have compensating advantages in its application to the measurement of sound-intensity under water. As yet, however, there is insufficient information regarding the water load and relative velocity-amplitude of such an arrangement to warrant its use as a standard of intensity-measurement.

§ 6. ACKNOWLEDGMENT

In conclusion the author desires to thank the Admiralty for permission to publish this paper.

REFERENCES

(1) König, W. *Ann. Phys.*, Lpz., **43**, 43 (1891).
(2) Crandall, L. B. *Vibrating Systems and Sound*, p. 137.
(3) Boyle, R. W. *Trans. roy. Soc. Can.* **20**, 233 (1926).
(4) Wood, A. B. and Smith, F. D. *Proc. phys. Soc.* **47**, 149 (1935).
(5) König, W. *Ann. Phys.*, Lpz., **42**, 549 (1891).

534.12

AN EXPERIMENTAL DETERMINATION OF THE FREQUENCIES OF FREE CIRCULAR PLATES

By A. B. WOOD, D.Sc., F.Inst.P.

Received March 15, 1935. *Read June* 7, 1935

ABSTRACT. The paper deals with an experimental determination of the frequency of vibration of circular plates of appreciable thickness. Two sets of observations made (i) with steel discs varying in diameter from 6 to 16 cm. and in thickness from 0·07 to 2·8 cm., and (ii) with discs of various materials all 2 cm. in diameter but varying in thickness, are described. Frequencies were measured at which the discs vibrated with (*a*) two nodal diameters, (*b*) one nodal circle. The experimental results are compared with theoretical values. The method may be applied to determine the velocity of sound in materials available only in small quantities.

---

§ 1. INTRODUCTION

The complete theory of the vibration of a free circular plate was developed by Kirchhoff in 1850. Before this (in 1829) Poisson had investigated theoretically some of the symmetrical modes of vibration and calculated the diameters of the nodal circles. Rayleigh$^{(1)}$ gives the general theory and quotes Kirchhoff's observations on the low frequencies of thin plates, and Lamb$^{(2)}$ summarizes the various theoretical results.

N It is stated that for a plate of known lateral dimensions the frequency N is given by

$$N=0{\cdot}0462\ ctm^2 \qquad \text{......(1)},$$

c, t where c is the velocity of sound appropriate to the material, t the thickness of the
m plate and m a constant for a plate of given dimensions. The lowest natural frequency of vibration is that in which the free circular plate has two nodal diameters, viz.

$$N=0{\cdot}261\ ct/a^2 \qquad \text{......(2)},$$

a where a is the radius of the plate. According to Poisson the mode of vibration having one circular node has a frequency N given by equation (1) in which

$$m^2=8{\cdot}89/a^2,$$

whence

$$N=0{\cdot}410\ ct/a^2 \qquad \text{......(3)}.$$

r The radius r of the nodal circle is given by

$$r=0{\cdot}678\ a \qquad \text{......(4)}.$$

It will be seen from relations (2) and (3) that the frequency varies directly as the velocity of sound and the thickness of the plate and inversely as the square of the radius.

So far as I am aware no information has been published relative to the vibrations of free circular plates of appreciable thickness. Such plates have high natural frequencies of transverse vibration which may lie outside the audible range. The present paper is mainly concerned with such plates. The experimental observations are conveniently divided into two series which deal with (i) circular steel plates ranging in diameter from 6 to 16 cm. and in thickness from 0·07 to 2·8 cm., and (ii) small circular plates 2 cm. in diameter and made of various materials.

§ 2. CIRCULAR STEEL PLATES

In this series of experiments the resonant frequency of the steel discs was determined by exciting them into vibration by means of a pot electro-magnet. The centre pole of this magnet, a bundle of soft iron wires, carried at its upper end the alternating-current winding, whilst a direct-current winding was provided in the usual manner. The steel discs rested in a horizontal position on a thin soft pad interposed between the disc and the face of the electromagnet. Sand was sprinkled over the surface of the disc and the frequency of the alternating current varied until the nodal diameters or nodal disc appeared. The tuning was found in all cases to be very sharp, so that the resonant frequency, determined by means of a calibrated wave-meter (or a monochord in the case of the lower frequencies), was accurately known. Observations in this series were confined almost entirely to the frequency at which the nodal circle appeared; in two cases only were the frequencies corresponding to two nodal diameters recorded. The more important results are included in table 1.

Table 1. Mild steel discs vibrating with one nodal circle

Number	Diameter $2a$ (cm.)	Thickness t (cm.)	t/d	Frequency (c./sec.)	Diameter $2r$ of nodal circle (cm.)	r/a
1	6·3	1·00	0·159	19,950	4·23	0·670
	6·3	1·00	0·159	(12,300—2 nodal diameters)		—
2	6·82	0·95	0·139	16,500	4·60	0·674
3*a*	7·00	1·45	0·207	21,800	4·65	0·665
	7·00	1·45	0·207	(13,750—2 nodal diameters)		—
3*b*	7·00	1·20	0·172	19,000	4·65	0·665
3*c*	7·00	1·00	0·143	16,400	4·65	0·665
3*d*	7·00	0·95	0·135	15,800	4·65	0·665
4*a*	8·23	1·07	0·130	12,900	—	—
4*b*	8·23	1·02	0·124	12,650	5·55	0·673
5	8·80	0·125	0·0143	1,480	5·9	0·670
6	8·90	1·50	0·169	14,800	5·95	0·670
7	8·90	0·0735	0·0083	845	6·0	0·675
8*a*	8·90	1·80	0·204	16,900	5·95	0·670
8*b*	8·90	1·50	0·169	14,800	5·95	0·670
8*c*	8·90	1·00	0·1125	10,650	6·0	0·675
8*d*	8·90	0·60	0·067	6,700	6·0	0·675
9*a*	15·5	2·77	0·178	8,750	10·4	0·670
9*b*	15·5	1·0	0·065	3,700	—	—

It will be seen in all cases that the ratio r/a of the radius of the nodal circle to the radius of the disc shows very little variation with the thickness and diameter of the discs. The extreme values are 0·665 and 0·675 whilst the theoretical value obtained by Poisson for a thin plate is 0·678. The experimental result is all the more remarkable when it is considered that the ratio of thickness to diameter of the discs in some cases reaches a value as high as 1 : 5.

With regard to the frequency, however, the results are not quite so simple. Figure 1 (*a*) shows the relation between thickness and frequency for discs of different diameters. The frequency increases at a rate which is slightly less than proportional to thickness. Similarly in figure 1 (*b*) the relation between frequency and $1/a^2$ is not quite linear. These results indicate a small but definite departure from the theoretical relationship indicated by equation (3) when the plates can no longer be regarded as thin.

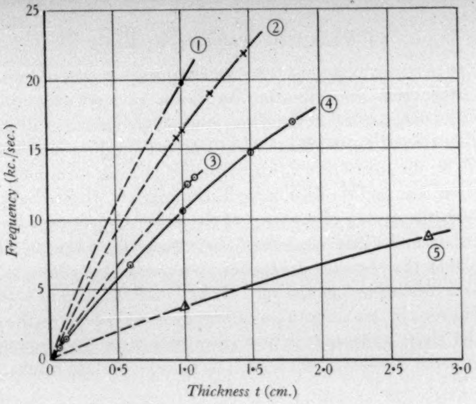


Figure 1 (*a*). Diameters ① 6·3 cm.; ② 7·0 cm.; ③ 8·23 cm.; ④ 8·90 cm.; ⑤ 15·5 cm.

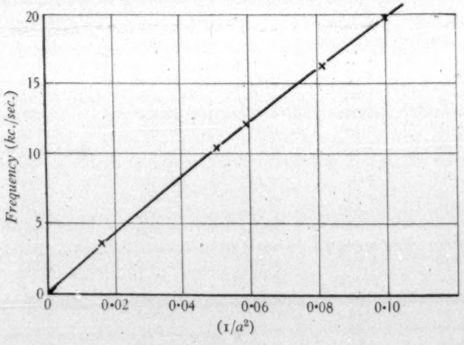


Figure 1 (*b*).

The observations for the thinnest plates, nos. 5 and 7 in table 1, lead to a value of $5{\cdot}6 \times 10^5$ cm./sec. for the velocity of sound in thin sheet steel, if the theoretical value 0·410 for the numerical constant in equation (3) is correct. On this basis the values of this constant *K* have been calculated for the thicker discs. The results are plotted in figure 2. The points all lie fairly well on a smooth curve in spite of the fact that the thickness and diameter of the discs are varied, independently, over a wide range. It will be seen that the value of *K* decreases fairly rapidly as the thickness of the disc increases, relatively to the diameter.

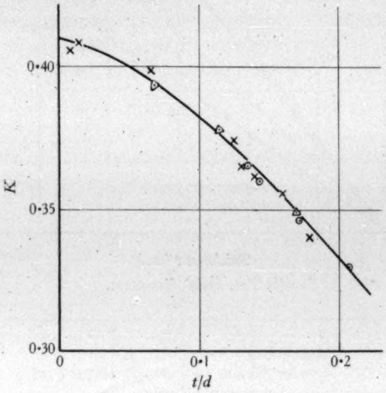


Figure 2.

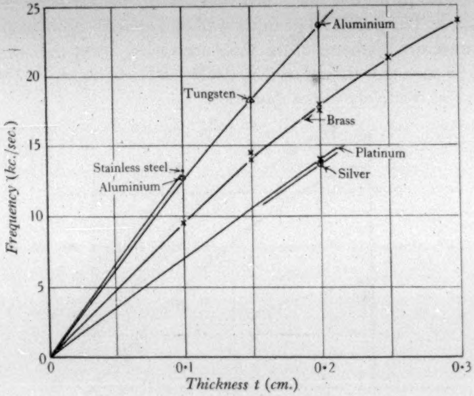
§ 3. OBSERVATIONS WITH SMALL CIRCULAR PLATES OF VARIOUS MATERIALS

In connection with the experiments described in the accompanying paper relating to measurement of sound-intensity in water by means of a Rayleigh disc, it was important to determine the natural frequencies of vibration of free discs 2 cm. in diameter. These discs were made of materials varying in density from 2·65 (aluminium) to 21·5 (platinum). The only available information relative to the frequency of vibration of such discs was of a purely theoretical nature, e.g. equations (2) and (3) above, and involved a knowledge of the velocity of sound in the material. Tabulated information bearing on this point was scanty and often unreliable. It was considered desirable therefore to determine the frequencies of such discs experimentally.

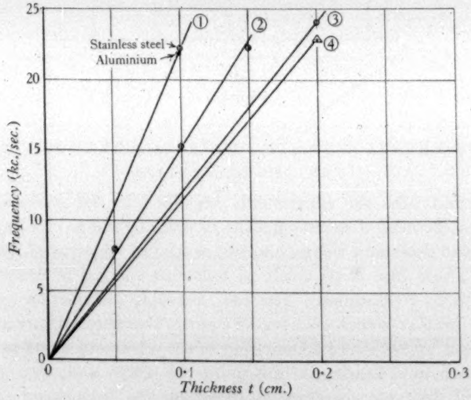
The method described in the previous section of the paper with reference to steel discs was inapplicable to this case, for most of the discs were non-magnetic. A simple and reliable method was devised in which the source of vibration of variable frequency was a small magnetostriction oscillator.* This consisted of a laminated nickel block supplied with a polarizing-winding carrying direct current, and an alternating-current winding connected to a valve oscillator having a wide range of frequency-control.

---

* As used in echo depth recording. See British patent specification No. 375,375.



(*a*) Two nodal diameters.



(*b*) One nodal circle.

Figure 3. Frequencies of metal discs (radius $a = 1$ cm.). ① aluminium and stainless steel; ② brass; ③ platinum; ④ silver.

The small disc under examination was placed on the horizontal vibrating surface of the magnetostriction oscillator and sprinkled with fine sand. The frequency of excitation was then adjusted until the required sand figure, two nodal diameters or one nodal circle, appeared. The frequency was then observed in the usual manner by means of a calibrated wave-meter. The results of such observations are shown plotted in figure 3 (*a*) and (*b*) where the frequencies, ranging from 8 kc./sec. to 23 kc./sec., are plotted as a function of thickness, (*a*) relating to two nodal diameters and (*b*) to one nodal circle. All the discs were 2 cm. in diameter.

As in the case of the large steel discs, the frequency increases at a rate which is slightly less than proportional to the thickness. The radius of the nodal circle was found to approximate closely to $0{\cdot}67a$. The dimensions and frequencies of each disc being known, the velocity of sound may be calculated from relations (2) and (3) above, the constants 0·261 and 0·410 being assumed. Velocities determined in this way and included in table 2 are therefore subject to a correction depending on the ratio of thickness to diameter as indicated in figure 2.

Table 2

Material	Velocity of sound (cm./sec.)	
	(*a*) from observations of two nodal diameters	(*b*) from observations of one nodal circle
Brass	$3{\cdot}6\times10^5$	$3{\cdot}7\times10^5$
Aluminium	4·9 ,,	5·4 ,,
Stainless steel	5·1 ,,	5·5 ,,
Silver	2·6 ,,	2·8 ,,
Tungsten	4·6 ,,	—
Platinum	2·7 ,,	2·9 ,,

There appears to be a tendency for the nodal-circle observations to yield somewhat higher values of velocity than those derived from the nodal diameters. Considering the quantity of material available in discs 2 cm. in diameter and 1 or 2 mm. in thickness, these experimental values for the velocity of sound are, however, in remarkably good agreement with those obtained in larger-scale determinations. The method shows promise therefore as a means of determining the velocity of sound in materials available only in small quantities, e.g. in the form of a small thin disc or button.

§ 4. ACKNOWLEDGMENT

In conclusion the author desires to thank the Admiralty for permission to publish this paper.

REFERENCES

(1) RAYLEIGH. *Sound*, 1, Chap. x.
(2) LAMB. *Dynamical Theory of Sound*, p. 152.

53.088.2

THE STATISTICAL THEORY OF ERRORS

By NORMAN CAMPBELL, Sc.D., F.Inst.P.

Received February 14, 1935. *Read in title June* 7, 1935

ABSTRACT. It is maintained that the adjustment of observations is *not* a statistical problem. In § 1 the problem is stated. In § 2 it is argued that there is no proof that the conditions necessary for a statistical theory of errors exist. In § 3 it is argued that the process of adjustment adopted by experimenters can be justified apart from statistical considerations. In § 4 the conclusions thus attained are reviewed. The actual argument cannot be summarized.

§ 1. INTRODUCTION

Deming and Birge⁽¹⁾ have recently discussed at great length the application of modern statistical methods to the theory of errors. They are probably not so sanguine as to believe that their work will have much effect on the practice of experimental physicists. It is as certain as anything can be that the great majority of us will fail to use the weapons they offer, and will continue to employ only the very simple methods of adjusting observations that are familiar to us; it is equally certain that this great majority will include almost all of the greatest masters of the experimental art.

And yet, if Deming and Birge's opening statement is true, we and our masters must be wrong. If "problems that arise in drawing conclusions from observations are essentially statistical", then it cannot be right to use for the solution of such problems any but the most perfect methods known to modern statisticians. If we are to justify ourselves, we must attack the statistical theory of errors much more fundamentally than heretofore; we must cease to object to particular statistical methods and assumptions, and boldly deny from the outset Deming and Birge's fundamental statement. We must deny that the adjustment of observations is a statistical problem at all.

A frivolous analogy may be useful. A golfer sees some men hitting a ball about with bent sticks. Noticing that they are urging the ball in a prescribed direction, and that their sticks are not unlike those shown in 18th-century golf pictures, he concludes that they are trying to play golf. He upbraids them for their ignorance of the modern art, and expounds to them the virtues of rubber-cored balls, steel-shafted drivers and a panoply of irons. They have no defence if they merely say that modern golf is over-elaborate, or that their implements are good enough for the very crude form of the game they are playing; to silence him they must reply that they are not playing golf at all, but hockey; they must point out that their game, though presenting certain superficial resemblances to golf, is so fundamentally different that his instructions are not wrong, but irrelevant.

§ 2. STATISTICS

Statistics, or at least all the statistical methods of Deming and Birge's paper, are directed to obtaining information about a collection[*] from a sample drawn from it. The problem is trivial unless the members of the collection are known to differ amongst themselves, so that members of the sample may also differ. In the adjustment of observations we are concerned with differing measurements; this is the fact that has led statisticians to believe that such adjustment is a statistical problem. But this belief is not justified unless (*a*) the differing observations are a sample of a collection, the sample and the collection being of the kind with which statistics is concerned; and (*b*) complete knowledge of the collection would solve or help to solve the problem of adjustment. In this section we shall consider (*a*), leaving (*b*) till later.

If measurements M to be adjusted are a sample of a collection, this collection can only be one of three things. It must be either N_1, N_2 or N_3, where $N_1 = M +$ all other similar measurements that are known; $N_2 = N_1 +$ all the similar measurements that will be known; and $N_3 = N_2 +$ all the similar measurements that might be known. If the collection is N_1, the problem can be, and ought to be, removed at once from the realm of statistics by discussing not the sample M but the whole collection N_1. If it is N_2 it can be, and ought to be, similarly removed by proceeding at once by making such a complete series of measurements that nobody will ever want to add to it. Accordingly the problem is essentially statistical only if the collection is N_3. This is apparently Deming and Birge's view. It does not seem to me at all sure that N_3 means anything at all; I cannot clearly conceive of a measurement that might be or might have been made and yet never will be or has been made. But the discussion may proceed on the assumption that the collection is N_3.

M
N_1
N_2
N_3

We have next to inquire what are the samples and collections to which statistical theory is applicable. One answer, which might have been given at the end of the 18th century, is that it is applicable to all samples of all collections. Nobody surely will give that answer to-day; everyone recognizes that it is absurd to treat statistically the answers of a jury in a criminal trial. The reasons why that answer is false have been given very fully in our own day by Keynes. But beyond that there is no agreement. Statistical methods depend on the conception of a measurable probability[†], and there is as yet no agreement as to what propositions have probability and how they come to have it. Progress has been made, for example by Mises and by Jeffreys, in determining what mathematical axioms must be asserted about probabilities in order that the calculus of probability may be applicable to them; but the difficulty is to relate these axioms to facts. How wide a difference of opinion still exists in this matter may be seen from any discussion between adherents and opponents of the so-called frequency theory of probability[‡].

---

[*] The usual term "population" is undesirable here, for it has too definite connotations.

[†] I use *probability* to denote a property of propositions, and distinguish it from *chance*, which is a property of material systems. See N. R. Campbell, *Phil. Mag.* **44**, 67 (1922).

[‡] See for instance the controversy between R. A. Fisher and H. Jeffreys, *Proc. roy. Soc.* A, **138**, 48 (1932); **139**, 343 (1933); **140**, 133 (1933); **146**, 1 (1934). The question which led to the controversy has been settled in the opinion of many qualified to judge; but its settlement leaves quite unreconciled the directly contradictory statements made concerning the basis of all statistical treatment.

However, there are collections of two sorts to which everyone admits statistics to be relevant; they are typified respectively by the urn containing black and white balls and the die or the tossed coin. The first sort of collection is finite and enumerable; it is possible for the proportion of members of the various kinds to be accurately known when the sample is drawn. The second sort is not finite or even definite; it is impossible that this proportion should ever be known accurately; but, on the other hand, very large samples reveal that there is one single proportion, or possibly a narrow range of proportions, definitely characteristic of the collection. Some would say that this proportion is the limit to which the proportion in the samples tends as their size is indefinitely decreased; others deny the validity of that conception and maintain that the proportion can be ascertained only with the help of *a priori* assumptions. But they all seem to agree that the large samples are necessary; by no other means can (e.g.) a true die be distinguished from a weighted die or from something that is not regulated by chance at all.

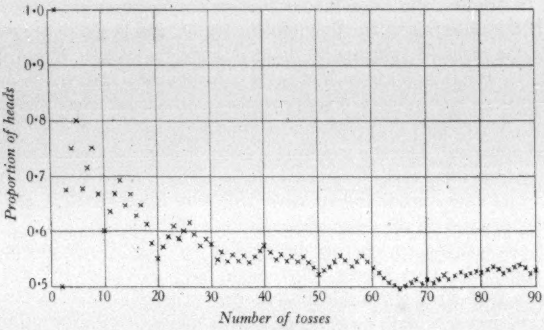


Figure 1.

The collection N_3 is clearly not of the first sort; it is not finite and enumerable. It may be of the second sort; but then we have no large sample. It must be insisted with the greatest possible emphasis that a single sample in which no regularity is evident does not afford sufficient evidence that the collection from which it is drawn is of the second sort and possesses the necessary characteristic proportion, or chance. On this point all serious statisticians seem agreed.

It is unusual to have as many as 20 measurements made in conditions so similar that, on any reasonable theory of error, they can be regarded as a sample drawn from a single collection. How inadequate such a sample is to prove the existence of a chance is illustrated by figures 1 and 2. These show the results of 500 tosses of a coin recorded for this occasion; the proportion of heads is plotted against the number of trials in the two diagrams on different scales. The first 20 tosses suggest no characteristic chance; the first 90 suggest one definitely greater than 0·5; the first 500 leave the whole matter doubtful. Many more tosses would be required to convince anyone that the chance really exists.

The application of statistics to adjustment cannot therefore be justified by a similarity between the measurements to be adjusted and a collection of either of these two typical sorts. Many statisticians would doubtless maintain vehemently that their methods are applicable to other sorts of collections and samples; but the arguments advanced for such application are very various and often mutually inconsistent. In particular it seems to me that the arguments for applying the conception of probability to other sorts of collections are often destructive of statistics in the sense in which I am using the word, namely the rules advanced by Deming and Birge for reaching conclusions concerning a collection from a sample of it. If those rules are valid in respect of our problem, no generally accepted proof of their validity has yet been offered.*

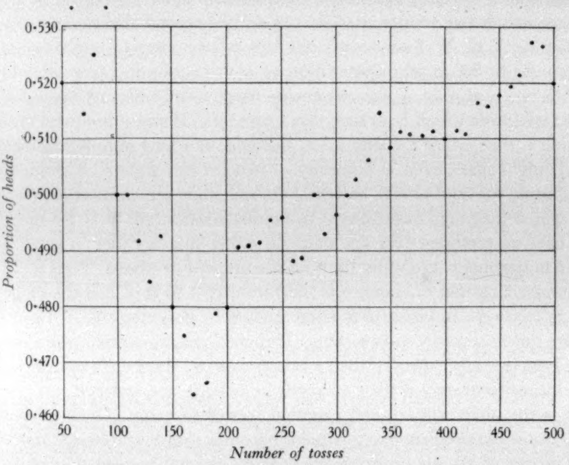


Figure 2.

Having thus found that (in terms of the analogy) it is by no means certain that the game that physicists are playing is statistics, let us inquire whether there is any other game, having definite rules, according to which their procedure can be justified. In the next section I shall try to state such rules.

---

* Those who think that it is easy to decide when and why statistical theory *is* applicable might study this question. Complete rainfall records are available for the years 1901–20; was the climate between 1901 and 1910 different from that between 1911 and 1920? If this problem is statistical, the records for 1901–10 and for 1911–20 must be samples of two possibly different collections. But what are the remainders of these collections? Not the records for other years; for, if the climate may be changing, other years are not comparable. But meteorological records must be records for some defined period. If the records for 1901–10 are a mere sample of the records for some longer period, and not the whole collection relevant to the problem, what is this longer period?

§ 3. THE ADJUSTMENT OF OBSERVATIONS

x

X_1

X_2

The simplest problem. Suppose that I am measuring a well-defined length x on a millimetre scale. I put one end against the zero and denote the position of the other end by a numeral X_1. Subsequently I repeat the process and record another numeral X_2. If X_1 is the same as X_2, no question of adjustment can arise. If X_1 and X_2 are different, there are two alternative assumptions. Either the length has changed, or one (or both) of the observations are in error. If the first alternative is adopted there is no justification for any adjustment, apart from certain circumstances to be discussed on page 806. I have measured two different things, and there is an end of it. Only the second needs examination.

α, β

The reason why we are prepared to consider the second alternative is this. When I am assigning X_1 or X_2, I am never quite sure what numeral I ought to assign. All that I can really do, at any observation of any magnitude except number, is to distinguish two ranges for X, namely a wide range α and a narrow range β within it. α is the range inside which I am sure that x must lie. If any subsequent observation lies outside α, then either the length has changed, or I have simply made a mistake, such as I might make even in counting. Mistakes can always be eliminated with sufficient care; we shall always assume here that they are eliminated; no observations in which they may occur ought to be discussed at all. β is the range within which I have no preference for one numeral rather than another; it is conveniently expressed in decimal notation by limiting the number of places. Thus if X_1 is 56·3, I am only asserting that the range α is (say) from 56·1 to 56·5, and the range β from (say) 56·25 to 56·35. In virtue of this uncertainty in assigning any X, two values X_1 and X_2 are not necessarily inconsistent. If their β ranges overlap, they are as consistent as I can possibly expect; if their α ranges overlap, these are possibly consistent; they are inconsistent only if their α ranges lie outside one another.

That is the only primary and essential source of error. Once we realize the nature of error, the problem of adjustment becomes simple and its solution obvious. If the β ranges of all our observations overlap, the assignment to x of any value lying in their common part will assert a scientific fact as definite as any fact can be. The selection of one of these values rather than another (e.g. the mean or median) is purely conventional and adds nothing to the fact. If the α ranges but not the β ranges overlap, then there is no definite scientific fact that can be asserted with full confidence; the possibility cannot be excluded that the magnitude is really different at different observations and therefore that adjustment is not permissible. If we decide to reject this possibility, there is some reason to accept the value that lies farthest from the boundary of any α range; for the subjective probability of a variation of the magnitude is greater, the closer is any value that we accept to the boundary of its α range. This preferred value usually approximates closely to the mean of the observations; there is then some reason for preferring that value, but even more for preferring the median.

These considerations determine also how many observations I should make. If I can get a few observations with overlapping β ranges, no increase of their number can possibly increase the certainty of any value that may be selected. If all further observations still overlap in their β ranges the field of selection may be narrowed, but not the certainty of the scientific fact. But if any observation should occur which overlaps only in α ranges, doubt is immediately introduced; if eventually a condition were reached in which only a very narrow range common to all α ranges survived, the range of selection would be very narrow; but the evidence for any value in that range would be practically nil. Accordingly no more observations should be taken than are sufficient to establish that values with overlapping ranges, α or β, can be obtained.

Curve-fitting. But pure science is never concerned with the value of a single magnitude; it is interested only in relations between magnitudes. The problem that we have just discussed is therefore of little importance, but the discussion enables us to deal briefly with important problems. Of these one will be taken as typical; the application of our treatment to the remainder will be obvious*.

In this problem we desire to know whether a numerical law

$$f(x, y) = a \qquad \text{......(1)}$$

is true. x, y are two different magnitudes, f is a known function defining the form of the law, a is a constant whose value is unknown. We measure n examples x_r ($r = 1$ to n) of the magnitude x, and the magnitude y_r corresponding to each of them. (The exact meaning of "corresponding" need not be discussed here.) For x_r we shall obtain a set of observed values X_{rs} ($s = 1$ to m_r) and for each y_r a set $Y_{rs'}$ ($s' = 1$ to m_r'). We try to find in the β ranges of each X_{rs} and each $Y_{rs'}$ values (X_{ro}, Y_{ro}), and also a value of a, such that

$$f(X_{ro}, Y_{ro}) = a \qquad \text{......(2)}$$

x, y, f

a, n, x_r

y_r

$X_{rs}, Y_{rs'}$

X_{ro}, Y_{ro}

for all values of r. If we succeed, then our previous discussion shows that we have all the evidence for the truth of equation (1) that we can have for any scientific fact. If we fail in this case but succeed when we use α ranges instead of β ranges, then the law is possible but not certain; it is the less probable the more we have to use values X_{ro}, Y_{ro} near the boundary of their α ranges. If we cannot find values in the α ranges, the law is certainly false.

The way in which this test is usually applied is slightly different in detail, but the same in principle. It is used because it is much more convenient, and indeed indispensable in all but the very simplest cases. We select a value of a by some rule or method, such as drawing a straight line, which requires only a knowledge of the X_{rs}, $Y_{rs'}$. We then find the residuals resulting from the substitution of the X_{rs}, $Y_{rs'}$ in equation (1), and compare these residuals with certain ranges calculated from the α and β ranges of the observations; it is unnecessary to set out here the well-known process of calculation. The three conclusions then follow according as the residuals all fall in the ranges calculated from the β ranges, in those calculated from the α ranges, or outside both.

The conclusions thus attained apply only to the law with the selected value of a. If the law does not turn out to be certainly true we ought, in all strictness, to repeat the test with other values of a. But there is one method of selection (or curve-fitting) of which it is commonly assumed that, if it does not give a value of a which makes the law true, no such value exists; this is the method of least squares.* I do not desire to throw any doubt on this assumption, which I believe to be true, but it must be insisted that the formal evidence for it is very slight; I have never seen a deliberate test made. The further assumption, often made, that the method of least squares is the only method of which this is true, and therefore that the value of a obtained by it is the only "right" value, rests on even less evidence. Indeed such evidence as exists is directly contrary; examples can easily be found in which the value obtained by other methods is so close to that obtained by the method of least squares that, if one is satisfactory, there can be no evidence against the other.

---

* Thus a relation between universal constants can be considered as a numerical law of which there is only one example.

These alternative methods have been proposed for the smaller amount of computation they involve and the simplicity of that which remains. The simplest is that known as zero sum; some others require even less arithmetic, but are a little more difficult to grasp in the first instance.†

Some of them are professedly, others actually, valid only in a limited class of cases; some selection among them may be needed. The advantage of the method of least squares—probably the only advantage—is that it can be legitimately applied without examination to any problem in which curve-fitting is permissible at all, that is to say, in which the form of the law given by f is definitely known.

Corrections. The very simple conditions discussed on page 805 hardly ever arise in practice; usually corrections have to be applied to the directly observed magnitude. Thus in measuring length, a correction for varying temperature may have to be applied. Now it is definitely known that we are not, or may not be, measuring the same magnitude at each observation; but adjustment is still permissible for the following reason. In applying the correction we are in effect assuming that there is some law

$$\xi = F(x, u, v, \ldots, a, b, c, \ldots) \qquad \ldots\ldots(2),$$

x, u, v
a, b, c
ξ

where x is the magnitude actually observed, $u, v, \ldots$ are other measurable magnitudes in virtue of which corrections must be applied, $a, b, c, \ldots$ are known constants, and ξ is the magnitude that we really want to determine. (Thus x may be measured length, u temperature, a a coefficient of expansion, b a standard temperature and ξ a length at the standard temperature.) We measure $x, u, v, \ldots$ and calculate ξ from equation (2). The measurements of $x, u, v, \ldots$ will all have their α and β ranges, unless one of them is number; and corresponding to their α and β ranges (in a sense that again need not be explained) will be α and β ranges for ξ. If we use these calculated ranges for the corrected magnitude, instead of the original ranges for the uncorrected magnitude, all our discussion of tests proceeds as before. The tests again are, in principle but not exactly in detail, those actually applied by the best experimenters in their most careful work.

Sometimes, perhaps usually, in addition to these known corrections there are unknown corrections. Thus we may know that the temperature of the rod is not uniform, but be unable to determine the distribution of temperature. If the error due to neglecting these corrections is not known to be limited, there is no justification for making any measurements at all. If it is known to be limited and the limit is known, the effect of neglecting the correction is to extend the α range. But there will not be in general anything corresponding to a β range. If the error is known to be systematic, i.e. always of one sign, there is certainly no β range; and even if there is no reason to suspect it to be of one sign rather than another, and if there is a range of complete and uniform uncertainty, it has not the same qualities as a true β range. For a true β range is inherent in the nature of measurement and of numerical laws; if there were not β ranges, no numerical law could be true.* An uncertainty due to an ignored correction is not of this character.

---

* By the "method of least squares" I mean an empirical rule for selecting one out of an infinitude of approximate solutions of a set of equations. I imply nothing about the theoretical assumptions by which the use of this rule is sometimes justified.

† See, for example, T. Smith, *Proc. phys. Soc.* **46**, 560 (1934) and J. H. Awbery, *ibid.* p. 574.

If then there are ignored corrections giving rise to an error comparable with the β range of the measurement, there can never be certainty either of an adopted value or of a numerical law. For only α ranges are present and any test based on them can never establish a certainty. In these circumstances it is best not to make any adjustment for the ignored correction. We should adjust as if it were absent, and then merely point out what degree of uncertainty attends the result.

§4. CONCLUSIONS

Such, I suggest, are the rules of adjusting observations as practised by the best experimenters.† There is necessarily some uncertainty attending the process. Accordingly those who believe that the calculus of probability is applicable to all propositions that are neither certainly true nor certainly false will hold that the calculus ought to be applied to adjustment. But here I am asking only whether the particular branch of the calculus concerned with collections and their samples is applicable; let us examine this question again, directing our attention now to condition (*b*) of page 801.

There are three kinds of collection which are, or may be, concerned in adjustment. First, if we are examining the truth of a law there is the collection of all examples of the law, which defines its form. Knowledge of this collection would be useful, but statistics does not profess to give it. The form of the law is always assumed to be known, and the problem is thus made substantially identical with that of determining a single value.

Secondly, there is the collection of all possible unknown corrections. But knowledge of this does not help us. We want to find out, not what corrections might be necessary (for we usually know that), but what corrections are necessary; in other words it is the sample, not the collection, that interests us. Further, it should be observed that, when (as is usual) the unknown corrections tend in one direction rather than the other, statisticians admit that they are powerless; for this is the case of systematic error.

---

* See, for example, my *Measurement and Calculation*, p. 153.

† As an example of this practice the paper by J. E. Sears, Jr., and H. Barrell, *Philos. Trans.* A, 233, 143, 216 (1934), may be quoted. But of course that is merely an example.

Thirdly, there is the collection of all values lying within the α (or β) range. Of this we do require information; for the collection defines the limits of the range. If our only source of information about the range were the recorded observations, then statistics would be useful—if it could determine the collection from the sample. But these are not the only source. The range arises from a certain indeterminacy in the judgment of the observer; we have, or should have, the testimony of the observer himself, which is the really decisive criterion. If the observer has been too careless or too incompetent to give us his opinion, his work is not worth considering at all. In any case it is surely the height of folly to deduce something about the judgment of an observer by means of an assumption that the observer himself would deny.

But such an assumption underlies all statistical treatment. Deming and Birge assume that "the effect of accidental errors can be reduced as far as desired by taking enough observations". If this means (as it must, if it is to be relevant) that the value of the measurements increases indefinitely with their number, the statement is one that every experimenter would deny. He limits the number of his observations, not because he is lazy, but because he knows that certainty is decreased, rather than increased, by increasing them beyond a quite small number. His reason for making more than a single observation is not to find out what is his α range; that he knows when he takes a single observation. It is to gain some assurance about unknown corrections, to ascertain that the actual spread of his observations is not so much greater than that arising from his α range that variable unknown corrections have to be postulated. He refuses to increase his observations beyond a certain point because he knows that thereby he increases enormously the probability of introducing new unknown corrections.

Herein lies the fundamental fallacy of the statistical theory of errors. It seeks to eliminate completely the judgment of the observer, which is an essential factor in determining the worth of his observations.* The plain fact is that if the observer is not capable of making the best of his measurements, nobody else can do it—except possibly by drawing his attention to mere mistakes that he will immediately recognize as mere mistakes.

To say this is not to deny great value to the work of those who collate the work of different experimenters. If they themselves are possessed, as Prof. Birge is, of experimental insight, they may well reach valid conclusions which each of these experimenters has missed. But the value of their work lies in their experimental insight and not in their statistical technique. If the elaborate methods of computation that Prof. Birge uses in his discussion of universal constants are replaced by the simple methods used by every physicist, while his judgments concerning the merits and interrelations of the various determinations are retained, everything of real value will survive. A neat series of best values, each with a tidy probable error, will not be obtained; but this series is positively misleading; it suggests something contrary to every principle of measurement, namely that there is a "true" value that can be found by experiments on a single magnitude. But his conclusions concerning the relations between the universal constants, such as that predicted by Eddington, will still be reached. Let anyone who doubts that carry through the work for himself, using nothing but arithmetic means and graphical curve-fitting.

---

* The absurdity reaches a climax when the results of different observers working in entirely different conditions are treated as a sample of a single statistical collection. But this is not an essential error. Here I am avoiding criticism of the erroneous, but unessential, assumptions of statisticians; but it may be worth while to draw attention to what seems to be a new error.

Deming and Birge assert (*loc. cit.* p. 124) that "the step...of the instrument...has the effect of grouping the observations into class intervals". If we are observing a single magnitude there is no plurality of values which the observations may classify, as the observations of (say) human stature may classify the statures of the individuals composing the population. Such classification by the measuring instrument in virtue of its "step" can arise only if the magnitude under observation is varying during observation. It may be—I think it is—meaningless to assert that there is a "true" value, and therefore meaningless to assert that the true value is constant; but then it is equally meaningless to assert, as apparently Deming and Birge do, that the true value is varying and assuming a plurality of values which the instrument classifies. In other words, there are involved in a measurement at most two things, the object we are trying to measure and our measurements of it. Deming and Birge seem to assume that there is a third thing intermediate between them; I cannot see what this third thing can be.

REFERENCE

(1) Deming, W. E. and Birge, R. T. *Rev. Mod. Phys.* **6**, 119 (1934).

621.385.2

THE VARIATION OF VOLTAGE-DISTRIBUTION AND OF ELECTRON TRANSIT-TIME WITH CURRENT IN THE PLANAR DIODE

By R. COCKBURN, B.Sc.

Communicated by Dr D. Owen, February 28, 1935. *Read in title June* 7, 1935

ABSTRACT. The general solution for the voltage-distribution in a diode with plane electrodes is derived from the differential equation of motion of the electrons, the effect of initial velocities being treated approximately. The solution is applicable for currents from zero up to that value which reduces the cathode field to zero. The general solution for electron transit-time is then derived. It is found that increase of space charge increases the transit-time, most of this increase occurring just before the cathode field is reduced to zero. The effect of initial velocities of emission is to reduce the transit-time, the amount of the reduction increasing as space charge increases, so that the sharp increase of transit-time due to the accumulation of space charge becomes less apparent for large initial velocities. The application of the theory to electron oscillations is discussed.

§ 1. INTRODUCTORY

In the production of electron oscillations by means of the thermionic triode valve, it is generally accepted that the periodic time of these oscillations is dependent on the transit-time of the electrons between the valve electrodes. A method of calculating the variation of transit-time with current in the grid-anode space, after saturation of the space has occurred, has been given by Gill(1), but the variation with current of the transit-time in the cathode-grid space has not previously been treated. In the present paper the voltage-distribution and the electron transit-time between plane electrodes have been worked out for any value of current between the electrodes, excluding the case when a potential-minimum is formed in front of the cathode. The effect of initial velocities of emission, which cannot be neglected, has been taken into account.

The problem of the transmission of electrons between plane electrodes was first examined by Langmuir(2), who obtained the differential equation of motion of the electrons. Child(3) had previously obtained the same equation and applied it to the motion of positive particles. Langmuir neglected initial velocities and gave a particular solution for the voltage-distribution applicable only to the one value of current which reduced the field at the cathode to zero. In a later paper Langmuir(4) obtained the voltage-distribution taking into account the initial velocity-distribution of the electrons. The rigorous solution which he gave is only possible when the current between the electrodes is sufficiently great to produce a potential-minimum in front of the cathode.

§ 2. INITIAL VELOCITIES OF EMISSION

To obtain a solution applicable for all values of current from zero up to that value which reduces the field at the cathode to zero, it has been necessary to make an assumption with regard to the initial velocities of the electrons emitted by the cathode. Whereas according to Maxwell's law they are emitted with all velocities from zero to infinity, it is here assumed that they are all emitted with an average velocity u_0 such that

u_0

$$u_0 = \left(\frac{\pi k T}{2m}\right)^{\frac{1}{2}},$$

where T is the absolute temperature, k is Boltzmann's constant, and m is the mass of the electron.

T, k, m

The error thus introduced in calculating the voltage-distribution is zero when the current flow is small, since in this case the voltage-distribution is independent of initial velocities; but as the current increases the error will become sensible. In the limiting case in which the current is sufficient to reduce the cathode field to zero, the approximate solution may be compared with Langmuir's rigorous solution(4). The error is found to be normally of the order of only 1 per cent; see equation (14*a*) *et seq.* For greater currents, for which the cathode field is negative and a potential-minimum is formed in front of the cathode, this assumption with regard to initial velocities is not valid, since an initial velocity-distribution is implicit in such a condition.

Likewise in the calculation of the transit-time, the equation obtained gives in effect the transit-time of an average electron. This differs, however, from the average transit-time by an amount which is negligible in practice.

§ 3. THE VOLTAGE-DISTRIBUTION

Let anode and cathode be assumed to be infinite parallel planes separated a distance x_a, and let the cathode be capable of emitting any given number of electrons at an average velocity u_0. Let the anode be maintained at a constant positive potential V_a, the cathode being at zero potential. Let the current-density be uniform and equal to i.

x_a
V_a
i

At a distance x from the cathode, let V be the potential, u the velocity of the electrons, n the number of electrons per cm.^{3}, and κ the dielectric constant of the medium and let m and e have their usual meanings.

V, u
n, κ
m, e

Then, following Langmuir, we may write

$$i = neu \qquad \text{......(1)},$$

$$\tfrac{1}{2} m u_0^2 + Ve = \tfrac{1}{2} m u^2 \qquad \text{......(2)},$$

$$\frac{d^2V}{dx^2} = \frac{4\pi}{\kappa} ne \qquad \text{......(3)}.$$

From equation (2)

$$u=\sqrt{\frac{2e}{m}}\,.\,(V+V_0)^{\frac{1}{2}},$$

V_0 where $$V_0 e\equiv\tfrac{1}{2}mu_0^2 \qquad \text{......(4).}$$

On substitution in equation (1)

$$ne=\sqrt{\frac{m}{2e}}\,.\,i\,(V+V_0)^{-\frac{1}{2}};$$

then from equation (3)

$$\frac{d^2V}{dx^2}=\frac{4\pi}{\kappa}\sqrt{\frac{m}{2e}}\,.\,i\,(V+V_0)^{-\frac{1}{2}} \qquad \text{......(5).}$$

The assumptions $V_0=0$ and $dV/dx_{x=0}=0$ give Langmuir's (2) particular solution of the equation, which may be put in the form

i_s $$i=i_s=\frac{\kappa}{9\pi}\sqrt{\frac{2e}{m}}\,.\,\frac{V_a^{\frac{3}{2}}}{x_a^2} \qquad \text{......(6).}$$

To obtain the general solution we multiply both sides of equation (5) by dV/dx and integrate. Thus

k $$\frac{dV}{dx}=\{4\alpha\,(V+V_0)^{\frac{1}{2}}+k\}^{\frac{1}{2}} \qquad \text{......(7),}$$

α where $$\alpha\equiv\left(\frac{4\pi}{\kappa}\sqrt{\frac{m}{2e}}\,.\,i\right) \qquad \text{......(8).}$$

To perform the second integration, put

y $$y^2=\{4\alpha\,(V+V_0)^{\frac{1}{2}}+k\};$$

A then $$x=\frac{1}{4\alpha^2}\left\{y\left(\frac{y^2}{3}-k\right)+A\right\}$$

$$=\frac{1}{4\alpha^2}\left[\{4\alpha\,(V+V_0)^{\frac{1}{2}}+k\}^{\frac{1}{2}}\,\{\tfrac{4}{3}\alpha\,(V+V_0)^{\frac{1}{2}}-\tfrac{2}{3}k\}+A\right] \qquad \text{......(9).}$$

Now $x=0$ when $V=0$,

$$\therefore\ A=-\{4\alpha V_0^{\frac{1}{2}}+k\}^{\frac{1}{2}}\,\{\tfrac{4}{3}\alpha V_0^{\frac{1}{2}}-\tfrac{2}{3}k\}.$$

By means of equations (6) and (8) express α in terms of i/i_s, thus

$$\alpha=\frac{4}{9}\,\frac{i}{i_s}\,\frac{V_a^{\frac{3}{2}}}{x_a^2} \qquad \text{......(10).}$$

k_1 The subsequent working is simplified by introducing the new constant k_1 defined by the relation

$$k=\left(\frac{4}{3}\,\frac{V_a}{x_a}\right)^2\left\{k_1^2-\frac{i}{i_s}\left(\frac{V_0}{V_a}\right)^{\frac{1}{2}}\right\} \qquad \text{......(11).}$$

Substituting these values in equation (9), we have

$$\frac{x}{x_a}\,.\left(\frac{i}{i_s}\right)^2=\left[\left(\frac{i}{i_s}\right)\left\{\left(\frac{V+V_0}{V_a}\right)^{\frac{1}{2}}-\left(\frac{V_0}{V_a}\right)^{\frac{1}{2}}\right\}+k_1^2\right]^{\frac{1}{2}}\left[\left(\frac{i}{i_s}\right)\left\{\left(\frac{V+V_0}{V_a}\right)^{\frac{1}{2}}+2\left(\frac{V_0}{V_a}\right)^{\frac{1}{2}}\right\}-2k_1^2\right]$$
$$-k_1\left\{3\left(\frac{i}{i_s}\right)\left(\frac{V_0}{V_a}\right)^{\frac{1}{2}}-2k_1^2\right\} \qquad \text{......(12).}$$

Equation (7) now becomes

$$\frac{dV}{dx}=\frac{4}{3}\frac{V_a}{x_a}\left[\left(\frac{i}{i_s}\right)\left\{\left(\frac{V+V_0}{V_a}\right)^{\frac{1}{2}}-\left(\frac{V_0}{V_a}\right)^{\frac{1}{2}}\right\}+k_1^2\right]^{\frac{1}{2}} \qquad \ldots\ldots(7a),$$

and when $V=0$ and $x=0$

$$\frac{dV}{dx_{x=0}}=\left(\frac{4}{3}\frac{V_a}{x_a}\right)k_1 \qquad \ldots\ldots(7b).$$

Thus k_1 is directly proportional to the field at the cathode.

Now since $V=V_a$ when $x=x_a$, k_1 must satisfy the following equation, obtained from equation (12)

$$\left(\frac{i}{i_s}\right)^2=\left[\left(\frac{i}{i_s}\right)\left\{\left(1+\frac{V_0}{V_a}\right)^{\frac{1}{2}}-\left(\frac{V_0}{V_a}\right)^{\frac{1}{2}}\right\}+k_1^2\right]^{\frac{1}{2}}\left[\left(\frac{i}{i_s}\right)\left\{\left(1+\frac{V_0}{V_a}\right)^{\frac{1}{2}}+2\left(\frac{V_0}{V_a}\right)^{\frac{1}{2}}\right\}-2k_1^2\right]$$
$$-k_1\left\{3\left(\frac{i}{i_s}\right)\left(\frac{V_0}{V_a}\right)^{\frac{1}{2}}-2k_1^2\right\} \qquad \ldots\ldots(13).$$

By rearranging and squaring we reduce this to

$$6\frac{i}{i_s}\left(\frac{V_0}{V_a}\right)^{\frac{1}{2}}k_1+3k_1^2-4k_1^3=\frac{i}{i_s}\left\{\left(1+\frac{V_0}{V_a}\right)^{\frac{3}{2}}+\left(\frac{V_0}{V_a}\right)^{\frac{1}{2}}\left(3-\frac{V_0}{V_a}\right)\right\}-\left(\frac{i}{i_s}\right)^2 \qquad \ldots(13a).$$

The parameter equation (13a) gives three values of k_1 for each value of i/i_s. One positive root is introduced during the squaring operation and does not satisfy equation (13), and one of the other two roots is negative for all values of i/i_s. This is physically untenable since $dV/dx_{x=0}$ cannot be negative within the limits of the problem. There is thus a unique value of k_1 for each value of i/i_s and V_0/V_a.

Tables 1, 2 and 3 show the variation of k_1 with i/i_s for three different values of V_0/V_a. The given value of V_0/V_a is substituted in equation (13a) and the cubic in k_1 is then solved for the various values of i/i_s by successive approximations.

Table 1. Variation of k_1 with i/i_s; $V_0/V_a=0$.

i/i_s	k_1
0·0	0·7500
0·1	0·7047
0·2	0·6575
0·3	0·6080
0·4	0·5557
0·5	0·5000
0·6	0·4398
0·7	0·3733
0·8	0·2972
0·9	0·2028
1·0	0·0

Table 2. Variation of k_1 with i/i_s; $V_0/V_a=0·001$.

i/i_s	k_1
0·0	0·7500
0·1	0·7071
0·2	0·6621
0·3	0·6154
0·4	0·5665
0·5	0·5155
0·6	0·4600
0·7	0·4005
0·8	0·3349
0·9	0·2598
1·0	0·1665
1·0963	0·0

Table 3. Variation of k_1 with i/i_s; $V_0/V_a=0·0049$.

i/i_s	k_1
0·0	0·7500
0·1	0·7093
0·2	0·6673
0·3	0·6232
0·4	0·5784
0·5	0·5357
0·6	0·4812
0·7	0·4282
0·8	0·3713
0·9	0·3091
1·0	0·2389
1·1	0·1549
1·217	0·0

It will be seen that when $i/i_s \to 0$, $k_1 \to \frac{3}{4}$ whatever the value of V_0/V_a. By substituting these values in equation (12) and expanding the contents of the first square bracket we obtain the relation between the potential V and the distance x from the cathode, in the case of a negligible current, as

$$V=\frac{V_a}{x_a}.x.$$

Again when $dV/dx_{x=0}=0$, $k_1=0$ from equation (7*b*), and thus from equation (13*a*)

$$\frac{i}{i_s}=\left\{\left(1+\frac{V_0}{V_a}\right)^{\frac{3}{2}}+\left(\frac{V_0}{V_a}\right)^{\frac{1}{2}}\left(3-\frac{V_0}{V_a}\right)\right\},$$

whence, on substitution for i_s, equation (6)

$$i=\frac{\kappa}{9\pi}\sqrt{\frac{2e}{m}}.\frac{V_a^{\frac{3}{2}}}{x_a^2}\left\{\left(1+\frac{V_0}{V_a}\right)^{\frac{3}{2}}+\left(\frac{V_0}{V_a}\right)^{\frac{1}{2}}\left(3-\frac{V_0}{V_a}\right)\right\} \qquad \text{......(14)}$$

$$=\frac{\kappa}{9\pi}\sqrt{\frac{2e}{m}}.\frac{V_a^{\frac{3}{2}}}{x_a^2}\left\{1+3\left(\frac{V_0}{V_a}\right)^{\frac{1}{2}}\right\} \text{ approximately} \qquad \text{......(14}a\text{)}.$$

η Now $V_0=\frac{m}{2e}u_0^2=\frac{\pi}{4}.\frac{k}{e}.T$. If we put η equal to eV_a/kT, equation (14*a*) becomes

$$i=\frac{\kappa}{9\pi}\sqrt{\frac{2e}{m}}.\frac{V_a^{\frac{3}{2}}}{x_a^2}\left\{1+\frac{2{\cdot}66}{\sqrt{\eta}}\right\} \qquad \text{......(14}b\text{)}.$$

This coincides with the approximate solution given by Langmuir ⁽⁴⁾, who showed that the current given by this expression differs by less than 0·3 per cent from the exact value when $dV/dx_{x=0}=0$; for the small values of V_0/V_a normally obtained, the current given by equation (14) will accordingly differ by about 1 per cent from the exact value. Thus the assumption that all the electrons are emitted with the same velocity instead of with distributed velocities causes an error of the order 1 per cent in the limiting case when $dV/dx_{x=0}=0$.

§ 4. THE ELECTRON TRANSIT-TIME

From equation (2) we have

$$u=\sqrt{\frac{2e}{m}}.(V+V_0)^{\frac{1}{2}},$$

if we write $V_0e=\frac{1}{2}mu_0^2$ as before.

T Now the transit-time $T=\int_0^{x_a}\frac{dx}{u}$

$$=\int_0^{x_a V_a}\sqrt{\frac{m}{2e}}.(V+V_0)^{-\frac{1}{2}}\,dx$$

and from equation (7) $\qquad dx=\dfrac{dV}{\{4\alpha\,(V+V_0)^{\frac{1}{2}}+k\}^{\frac{1}{2}}}$;

whence integrating and writing as before

$$\alpha=\frac{4}{9}\frac{i}{i_s}\frac{V_a^{\frac{3}{2}}}{x_a^2}, \quad \text{and} \quad k=\left(\frac{4}{3}\frac{V_a}{x_a}\right)^2\left\{k_1^2-\frac{i}{i_s}\left(\frac{V_0}{V_a}\right)^{\frac{1}{2}}\right\},$$

we have

$$T=3\sqrt{\frac{m}{2e}}.\frac{x_a}{V_a^{\frac{1}{2}}}\frac{i_s}{i}\left[\left\{\left(\frac{i}{i_s}\right)\left\{\left(1+\frac{V_0}{V_a}\right)^{\frac{1}{2}}-\left(\frac{V_0}{V_a}\right)\right\}^{\frac{1}{2}}+k_1^2\right\}^{\frac{1}{2}}-k_1\right] \qquad \text{......(15)}.$$

It will be seen that when $i/i_s \to 0$ and $k_1 \to \frac{3}{4}$

$$T=T_0=2\sqrt{\frac{m}{2e}}\cdot\frac{x_a}{V_a^{\frac{1}{2}}}\left\{\left(1+\frac{V_0}{V_a}\right)^{\frac{1}{2}}-\left(\frac{V_0}{V_a}\right)^{\frac{1}{2}}\right\} \qquad \text{......(16)},$$ T_0

and when $k_1=0$ and $\quad \dfrac{i}{i_s}=\left\{\left(1+\dfrac{V_0}{V_a}\right)^{\frac{3}{2}}+\left(\dfrac{V_0}{V_a}\right)^{\frac{1}{2}}\left(3-\dfrac{V_0}{V_a}\right)\right\}$

$$T=T_s=3\sqrt{\frac{m}{2e}}\cdot\frac{x_a}{V_a^{\frac{1}{2}}}\,\frac{1}{(1+V_0/V_a)^{\frac{1}{2}}+2\,(V_0/V_a)^{\frac{1}{2}}} \qquad \text{......(17)}.$$ T_s

Equation (16) shows the effect of initial velocities on the transit-time when the current is negligibly small, and equation (17) shows the effect of initial velocities when space charge has reduced the field at the cathode to zero. Equation (15) with the parameter equation (13 *a*) gives the value of the transit-time for any value of current, from zero up to that value which reduces the cathode field to zero, and for any value of emission velocity. It includes equations (16) and (17), of course, as special cases.

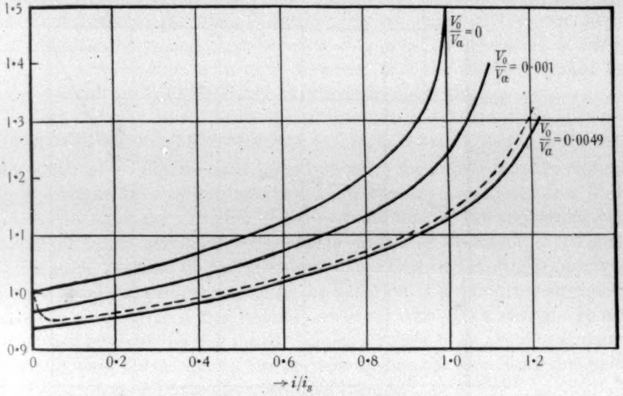


Figure 1. Variation of transit-time with current.

Figure 1 shows the calculated variation of transit-time T with degree of saturation i/i_s of the diode for different values of V_0/V_a. The ordinates are values of T/T_1 where $T_1=2\sqrt{(m/2e)}\,.\,x_a/V_a^{\frac{1}{2}}$. T_1

§5. APPLICATION TO EXPERIMENTAL CONDITIONS

In practice V_0/V_a varies with i/i_s, since both vary with the cathode-temperature, the emission velocity V_0 directly, and the current i nearly exponentially. Hence as i is increased from zero by raising the cathode-temperature, V_0 will increase rapidly at first and then quite slowly until $dV/dx_{x=0}=0$. The current is then limited by space charge and a very large increase in temperature, and hence of V_0, produces only a small increase in i. The type of variation of V_0 with i to be expected is illustrated in figure 2.

These results were obtained from measurements on a valve of A.T. 40 type with a tungsten filament. The ratio of amperes of emission current to watts dissipated in such a filament varies only with its temperature, and the variation in the case of tungsten has been given by Stead (5). From his results the approximate temperature of the valve filament for any voltage applied across it was obtained and the emission velocity was calculated.*

In figure 2 the abrupt change in slope to be expected at $dV/dx_{x=0}=0$ is rounded off owing to the velocity-distribution of the electrons, but it may be assumed that $dV/dx_{x=0}=0$ at point E, obtained by producing AB and DC to meet. Knowing now the value of V_0/V_a at E, we can calculate the value of i/i_s at this point from equation (14).

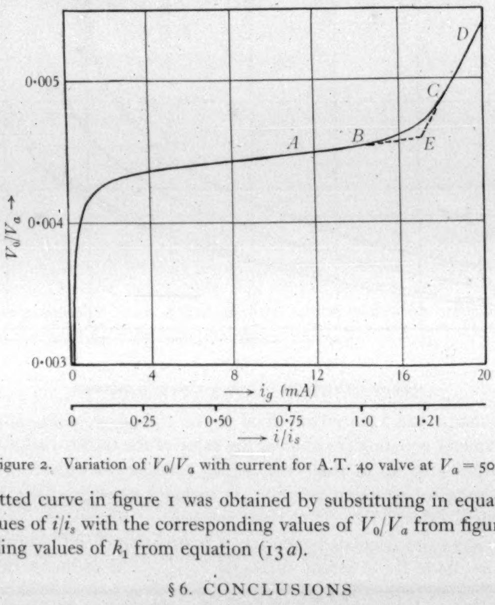


Figure 2. Variation of V_0/V_a with current for A.T. 40 valve at $V_a = 50$ V.

The dotted curve in figure 1 was obtained by substituting in equation (15) for various values of i/i_s with the corresponding values of V_0/V_a from figure 2, and the corresponding values of k_1 from equation (13 a).

§ 6. CONCLUSIONS

As the current between the plane electrodes increases, the space charge being consequently increased, the transit-time of electrons between the electrodes also increases, and most of the increase occurs just before the cathode field is reduced to zero. The effect of initial velocities of emission is to reduce the transit-time, the amount of the reduction increasing as space charge increases, so that the sharp increase of transit-time due to the accumulation of space charge becomes less apparent for large values of V_0/V_a. The effect of space charge on the electron transit-time should be most easily observed, therefore, with a low-temperature cathode and a large anode voltage.

---

* In the case of a cylindrical cathode $V_0=\frac{3}{2}.\frac{\pi}{4}.\frac{k}{e}.T$.

The theory has been worked out for plane electrodes, whereas in practice valves with cylindrical electrodes are normally used. Neglecting the effect of initial velocities, Schiebe (6) has calculated the electron transit-time between cylindrical electrodes when space charge is negligible, and McPetrie (7) has dealt with the case when space charge has reduced the cathode field to zero. The increase of electron transit-time due to space charge is of the same order as for plane electrodes. In another paper McPetrie (8) has extended Schiebe's treatment to include the effect of initial velocities on the transit-time when space charge is negligible. For the usual internal cathode the effect is considerably less than for plane electrodes. A general solution for transit-time between cylindrical electrodes is probably impossible. It is reasonable to assume, however, that the results obtained for plane electrodes will apply qualitatively to cylindrical electrodes.

§7. APPLICATION TO THE THEORY OF ELECTRON OSCILLATIONS

In the course of an experimental investigation by the author into the electron oscillations produced by a cylindrical-electrode triode valve connected in a Barkhausen-Kurz (9) circuit, the variation of the periodic time of the oscillations with grid current is being examined. For valves with high-temperature tungsten cathodes the usual decrease of periodic time with increasing current is observed. This effect is usually explained as due to the variation of electron transit-time in the grid-anode space after it has become saturated with space charge. It may also be explained by assuming that the oscillations are those of electrons in statistical equilibrium between the electrodes, the decrease in periodic time being due to the increase of concentration of the electrons with current.

In the case of one valve, however, with a low-temperature oxide-coated cathode, the decrease in periodic time was followed by a small but quite definite increase as the cathode-grid space became saturated with space charge. This increase was most evident at high grid voltages. These preliminary observations indicate definitely that the periodic time of the electron oscillations is determined by electron transit-time and that the path of these electrons is partly in the cathode-grid space. The investigation is proceeding with a view to establishing the complete path of those electrons which determine the periodic time of the oscillations.

REFERENCES

(1) E. W. B. Gill. *Phil. Mag.* **12**, 843 (1931).
(2) I. Langmuir. *Phys. Rev.* **2**, 450 (1913).
(3) C. D. Child. *Phys. Rev.* **32**, 498 (1911).
(4) I. Langmuir. *Phys. Rev.* **21**, 419 (1923).
(5) G. O. Stead. *J. Instn elect. Engrs*, **58**, 107 (1920).
(6) A. Schiebe. *Ann. Phys.*, Lpz., **73**, 54 (1924).
(7) J. S. McPetrie. *Phil. Mag.* **16**, 284 (1933).
(8) J. S. McPetrie. *Phil. Mag.* **16**, 544 (1933).
(9) H. Barkhausen and K. Kurz. *Phys. Z.* **21**, 1 (1920).

541·133

THE ELECTRICAL CONDUCTIVITY OF SOME STRONG ELECTROLYTES IN DILUTE SOLUTION AND ITS VARIATION OVER THE TEMPERATURE RANGE 18° C. TO 85° C.

By C. J. B. CLEWS, B.Sc., A.Inst.P., Queen Mary College, University of London

Received February 28, 1935. *Read in title June* 7, 1935

ABSTRACT. The work described in a previous paper has been extended to the measurement of the electrical conductivity of 0·0005-*N* solutions of potassium nitrate, potassium sulphate and sodium sulphate. These data, in conjunction with those previously obtained, provide strong evidence for the validity of the theory of Debye, Hückel and Onsager for uni-univalent electrolytes in dilute solutions up to 85° C. Agreement is not as satisfactory for sodium sulphate, while an anomaly has been observed in the behaviour of potassium-sulphate solutions.

§ 1. INTRODUCTION

In a previous communication to the Society (1) some measurements of the electrical conductivity of aqueous solutions of potassium nitrate, potassium sulphate and sodium sulphate were reported and data were given for solutions having concentrations down to *N*/1000. Owing to the Debye-Hückel-Onsager theory being applicable only in very dilute solutions, no deductions could be made as to the agreement of experiment with this theory in the case of uni-bivalent electrolytes, and only tentative conclusions could be made for potassium-nitrate solutions. It was in view of this difficulty that the present measurements were made.

§ 2. EXPERIMENTAL

The conductance has been determined with the same apparatus as before, with the exception that a new variable audio-frequency oscillator was used.

In extending the range to the more dilute solutions many difficulties were encountered. Among these was contamination of the conductivity water due very largely to working in a dirt-laden city atmosphere. There was also some difficulty in making up such dilute solutions with the required accuracy.

Owing to these and other minor troubles, it was only possible to measure the conductance at one more concentration, namely, *N*/2000, for the three electrolytes.

§ 3. RESULTS

The further data are given in table 1. Comparison with the work of other experimenters cannot be extensive, but where it is possible, i.e. 18° and 25°, agreement is quite satisfactory.

Table 1. Equivalent conductance of *N*/2000 solutions

Temperature (° C.)	Potassium nitrate	Potassium sulphate	Sodium sulphate
18	124·4	128·5	107·5
25	142·5	150·2	125·0
30	156·8	167·8	139·6
35	171·2	187·5	153·2
40	186·6	204·8	169·3
45	202·3	224·1	185·5
50	218·0	242·3	203·2
55	235·5	260·8	217·8
60	251·9	278·5	234·3
65	268·0	297·7	253·1
70	284·8	318·1	268·8
75	302·0	337·4	287·6
80	318·8	356·3	307·3
85	335·5	375·6	323·7

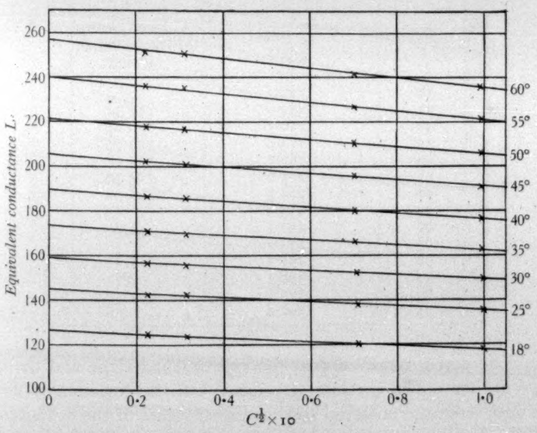


Figure 1. Potassium nitrate.

From the Debye-Hückel-Onsager theory we expect that a plot of the equivalent conductance against the square root of the concentration, at a given temperature, will yield a straight line; by this means we may readily test the theory. Figures 1 and 2 show how striking is the agreement for potassium-nitrate solutions, and furthermore, the values of the conductance at infinite dilution, L_0, given by the intercepts of these lines with the L axis are almost identical with those calculated from the theory, table 2*.

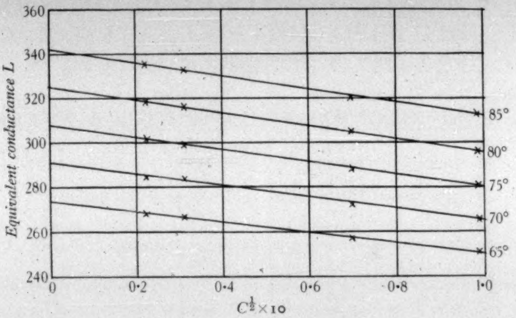


Figure 2. Potassium nitrate.

Table 2. Potassium nitrate. Equivalent conductance at infinite dilution

Temperature (° C.)	L_0 from graph	L_0 from Debye-Hückel-Onsager theory
18	126·3	—
25	144·6	144·54
30	159·1	159·04
35	173·8	173·78
40	189·7	189·68
45	205·6	205·39
50	221·5	221·30
55	239·5	—
60	257·1	—
65	274·2	—
70	291·3	—
75	308·1	—
80	325·7	—
85	342·4	—

It should be noted, however, that the experimental slope and the calculated slope $(\alpha L_0 + \beta)$† are not in absolute agreement; but the deviation is not serious, as may be seen from the observed and computed values of the conductance of an $N/100$ solution, table 3.

The results for potassium-sulphate solutions do not give us a linear graph, but give a curve which approximates to a parabola, except at the lower temperatures, figures 3 and 4. The curvature increases with the temperature. We are therefore led to the conclusion that even in fairly dilute solutions potassium sulphate does not behave as an ideal Debye electrolyte.

---

* Cf. reference (1), p. 770.

† α and β are factors involved in the Debye-Hückel-Onsager theory (*loc. cit.* p. 768).

Table 3. Potassium nitrate

Temperature (° C.)	Debye slope	Experimental slope	*L*, calculated *N*/100	*L*, experimental *N*/100
25	92·9	89	135·3	135·7
30	103·8	97	148·7	149·4
35	115·2	112	162·3	162·6
40	127·3	131	177·0	176·6
45	139·6	145	191·6	191·1
50	152·0	154	206·3	206·1

Some explanation may be offered by the Bjerrum theory (2) of ionic association. For uni-bivalent electrolytes the critical distance of approach is probably of the order 7 A., whereas the sum of the ionic radii of potassium sulphate is 3·7 A., so that we might expect strong association. This is, however, nothing more than a tentative and qualitative explanation.

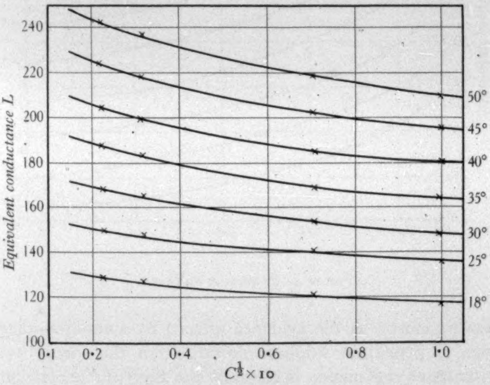


Figure 3. Potassium sulphate.

Contrary to expectations, sodium sulphate does not give parabolic curves; nevertheless, the linear plot is not as good as for the uni-univalent electrolyte.

The difference in the behaviour of these two sulphates provides a problem which may be rather difficult of solution, as it is usually found that similar groups of salts have very similar properties.

§ 4. CONCLUSION

The observations on potassium-nitrate solutions add definite support to the Debye-Hückel-Onsager theory, and particularly to its validity at higher temperatures. This is an extremely stringent test of the theory in that over the temperature-range 18° to 85° there is about a threefold increase in the conductance. We have already had ample proof, by Hartley ⁽³⁾ and his co-workers at Oxford, of the accordance of theory with experiment for dilute uni-univalent electrolytes at 25° in both aqueous and non-aqueous solvents. There can now be little doubt as to the fundamental truth of the Debye theory, particularly for uni-univalent electrolytes.

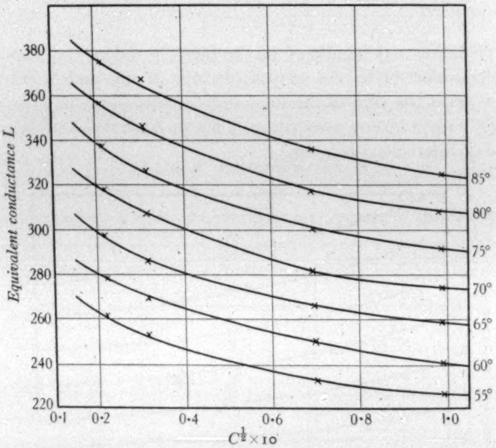


Figure 4. Potassium sulphate.

Rather less conclusive is the evidence offered by sodium-sulphate solutions, while solutions of potassium sulphate deviate from the theory even in dilute solutions. A tentative explanation is given by the Bjerrum association theory, but this is nothing more than qualitative.

That further information is required with regard to the apparently anomalous behaviour of potassium sulphate is obvious. Owing to local conditions the author will be unable to continue the investigations, and we can only hope that the necessary data, both on the behaviour of potassium-sulphate solutions and on the variation of conductance with temperature, may be forthcoming in the near future.

§ 5. ACKNOWLEDGMENTS

The work has been carried out in the Physical Laboratories at Queen Mary College, and my best thanks are tendered to Prof. H. R. Robinson, F.R.S., for the facilities he has placed at my disposal and for his advice throughout the investigation.

I am indebted to the Department of Scientific and Industrial Research for a grant which enabled me to carry out this work.

REFERENCES

(1) Clews, C. J. B. *Proc. phys. Soc.* **46**, 764 (1934).
(2) Bjerrum, D. *Kg. Dansk. Vid. Selsk. Medd.* **7**, No. 9 (1926).
(3) Hartley, H. *Proc. roy. Soc.*, A **109**, 351 (1925); **127**, 288 (1930); **132**, 427 (1931) etc.

537·56.06

AN IMPROVED COUNTING CIRCUIT

By H. P. BARASCH, Cavendish Laboratory, Cambridge

Communicated by C. D. Ellis, May 28, 1935. *Read June* 7, 1935

ABSTRACT. This paper deals with a simple Geiger-Müller counter circuit, the essential feature of which is the application of a special characteristic curve of a pentode valve. This $\{i_a, e_g\}$ curve has a saturation portion in the negative grid region. The output pulses from the circuit are equalized, free from secondary disturbances, and of shortened duration. By means of an appropriate resistance-capacity coupling to the thyratrons a back pulse is generated, so that the final pulse-shape is more suitable for triggering the thyratrons. A description of the simple pentode circuit is given and also of a universal set, consisting of four valves, which is suitable for coincidence counting. The special characteristic curve when applied to the coincidence circuit leads to improved discrimination between total and partial coincidences.

§ 1. INTRODUCTION

When a Geiger-Müller counter coupled to an amplifying circuit is in use, irregularities are sometimes observed whose cause is to be found not in the nature of the discharge itself, but rather in the circuits between the counter and the recorder. Closer consideration proves that even with correct discharges in the counter, double pulses which affect the recording circuit may be generated in the amplifier.

A good amplifier should have the following qualities: (i) sufficient amplification to actuate the recording device; (ii) good resolution—here a distinction is to be made between normal counting of a few hundred a minute and high-speed counting of a few thousand a minute; (iii) limitation, by which is meant the equalization of the pulse-amplitude; (iv) a clean, uniform output pulse-shape to prevent disturbances and false counts. For coincidence work there are two further points, which must be considered, viz. the provision of (v) high resolving-power for coincidences of the different counters, and (vi) good discrimination between coincidences of n counters and coincidences of $(n-1)$ counters.

In the last few years many amplifying-circuits for Geiger-Müller counters have been described. They usually involve several stages of amplification of the resistance-capacity-coupled type, and for coincidence counting still other stages are added. For example, in the circuit devised by Johnson and Street(1) three stages are necessary to fulfil conditions (i), (ii) and (iii). The first two stages amplify and equalize the pulses, the third reverses them and shortens their duration. In this way good resolution for coincidence work is obtained but further amplification is necessary because in the shortening of the duration of the pulses their size has been reduced. Parallel chains, according to the number of counters used, are necessary for coincidence counting.

An amplifier of the Johnson and Street type undoubtedly satisfies the conditions specified, but in simpler circuits condition (iv) is rarely realized. This demands the generation of a sharp uniform pulse-shape without wobbling, and it is not easy to prevent secondary potential disturbances in the coupling elements due to variation in size and form of the discharge in the counter itself. In particular the shortening stage, because of its low input time-constant, may generate a rather strong back-pulse, which is amplified by the valve and may give rise to a second forward pulse in the succeeding stages.

Visual observations on a cathode-ray oscillograph screen show that the output pulses from certain multistage amplifiers of the ordinary type not only are distorted and of long duration, but are followed by a series of undesirable potential-variations.

This paper will describe a method of using a single pentode, which in itself satisfies all the necessary conditions as a three-stage amplifier and also produces a shape of output pulse particularly favourable for thyratron recording.

§ 2. COUNTER DISCHARGE

A comprehensive treatment of the action of a Geiger-Müller counter has recently been given by Werner(2). Certain points in his arguments are relevant to the present problem, since they enable us to see in a general way the form of the pulse which is actually fed into the amplifying system, and how its shape and duration depend on the details of the coupling.

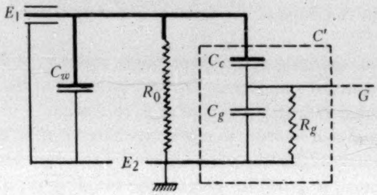


Figure 1.

When an ionizing particle or quantum ionizes the gas in the counter, the anode wire acquires in a very short time some negative potential, depending on the over-voltage applied to the counter. There it remains in a region of unstable corona discharge for some time t_m until the discharge stops abruptly. After this extinction, t_m the wire returns to zero potential in a time determined by the net capacity and resistances associated with the counter.

The two factors, the time taken for the potential to rise and the time during which the discharge burns in the counter, have to be considered in connection with the amplifying-circuit. One would expect that the former will determine the steepness of the pulse which arrives on the grid of the first valve. Werner has calculated the time t taken to charge the wire negatively to a small over-voltage and has ob- t tained a value of the order of $2 . 10^{-7}$ seconds(2).

An equivalent scheme of the counter circuit is shown in figure 1. The counter is represented by an e.m.f. E_1 with internal resistance K. C_w represents the wire- E_1, K, C_w earth capacity including all direct connections. C_c is the coupling-condenser and C_c C_g the effective grid capacity of the first valve with grid-leak resistance R_g. E_2 C_g, R_g, E

denotes the high-potential supply, assumed to have negligible internal resistance, and R_0 is the high resistance which serves to discharge the wire after extinction. K, the internal resistance of the counter, is a function of gas pressure, of the kind of gas used, and of the geometrical counter constants. At small values of the over-voltage it may be obtained from formulae or from tables given in Werner's paper. For a counter filled with air at a pressure of 8 cm. of mercury, K is of the order of 1 megohm.

During the charging process, when a particle enters there is a current flow i equal to $-C_0 dV'/dt$, where V' is the over-voltage applied to the counter. If R_0 is assumed to be so large that it does not affect the charging process, then C_0, the total earth capacity of the wire, is given by $C_0 = C_w + C'$, where C' represents the equivalent capacity of the impedance within the dotted rectangle. If we consider the pulse to be defined by the fundamental frequency ω we find that, approximately,

i, V'
C_0
C'

$$C' = \frac{C_c C_g (1 + 1/\omega^2 C_g^2 R_g^2)}{C_c + C_g (1 + 1/\omega^2 C_g^2 R_g^2)}.$$

Evaluating this formula with actual values of constants used and the fundamental frequency as given by the slope of the primary charging current to the wire, we find that $C_0 \cong 10$ to $15\,\mu\mu$F.

The time KC_0 of charging is correspondingly about $1 \,.\, 10^{-5}$ sec., a time much longer than that obtained if no external load is connected to the counter, for which we have already quoted Werner's estimate of $2 \,.\, 10^{-7}$ sec.

It is to be observed that in order to make the charging-time short it is essentially the direct earth capacity C_w of the wire, shown by the dark lines in figure 1, that should be made as small as possible, whereas the coupling-capacity C_c has a rather indirect effect, because of its position in series with C_g and R_g. The very small capacities involved render it necessary to avoid all superfluous metal parts connected to the counter wire which could increase the direct earth capacity. For this reason it was found convenient to make the coupling-condenser C_c by enclosing an extension of the counter wire in glass tubing of small diameter and wrapping metal foil around the outside of the glass. The capacity between the wire and the foil can be made about 20 to $40\,\mu\mu$F. without directly increasing the earth capacity.

In order that the wire potential after extinction shall come back to zero in a short time, so that the next pulse falling within the recovery time $R_0 C_0$ is large enough for amplification, it is necessary to make R_0 small. But a small R_0 means that the negative rise of wire-potential, governed by K, will be partially offset by charge from E_2 through R_0. This delays the rise of the wire-potential to the starting potential value and also forces the discharge to proceed to burn in a region of instable corona discharge which is more remote from the actual starting potential. In this region t_m is increased and the counter tends to continuous discharge. Too great a value of over-voltage has a similar effect. It is necessary therefore to have R_0 at least 100 times as great as K. A compromise must be made between a correct working of the counter and a reasonable resolving-power of the amplifier.

§ 3. PENTODE CHARACTERISTIC

The special pentode circuit we are discussing is shown in figure 2. The characteristic feature is that in series with the anode is a high resistance of about 250,000 Ω., so that the anode potential is about 3 V. The screen grid is kept at from 30 to 40 V. Under these conditions the characteristic shown in figure 3 (*a*) is obtained. It resembles the usual saturation curve of a hot emitter, but is unique in that the grid-voltage change required to go from cut-off to saturation of the anode current can be of the order of one volt. This range of voltage will be referred to henceforth as the *range of limitation* and denoted by L, figure 3 (*b*). L

The curve follows at first the usual slope of a pentode characteristic, but at a certain point it bends over very sharply into a horizontal line which extends from about -1 V. to high positive values. It is outside the scope of this paper to give an exact physical explanation of this saturation portion, but it may be stated briefly, that it is due to a space charge being established near the plate when the plate potential is maintained somewhat below 5 V.(3).

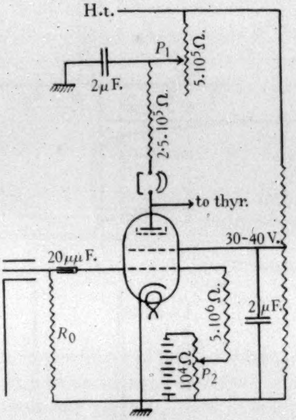


Figure 2.

In the present case it is the external resistance R_a which lowers the plate potential to about 3 V. A further consequence of importance is that the pentode valve in this region has an internal resistance of the very low order of about 5000 Ω. The behaviour in that state is analogous to that of a lighted thyratron, where the plate current is determined only by the anode resistance R_a and anode voltage V_a. The shape of the curve remains unchanged over a wide range of R_a and V_a. When R_a is decreased and V_a increased the current in the saturation part increases, but at the same time L increases. Satisfactory values found for a SP4 pentode valve are: h.t. voltage 150 V., $R_a = 250{,}000$ Ω. and screen voltage 30 to 40 V. The anode current i_a is then 0·63 mA. and $L = 1{\cdot}3$ V. R_a V_a

The working point A of the pentode is adjusted by means of a potentiometer to some point in the saturation region depending on the magnitude of the pulses on the grid. If this is sufficiently large, it is of advantage to shift A towards zero grid bias, because this gives an appreciable improvement in the shape of the pulse.

By the aid of this $\{i_a, e_g\}$ characteristic curve the conditions set out in the introduction are satisfactorily fulfilled. The adjustment in the saturation region corresponds to a bias which prevents not only the small backpulses but also all disturbances from being amplified; the latter may take the form of a.-c. and d.-c. hum, potential-variations due to thermal agitation in the high-input resistances, and general pick-up. Because of the small amplitude of these disturbances at the first grid, the bias (that is, the shift into the saturation region) need not be greater than a small fraction of a volt.

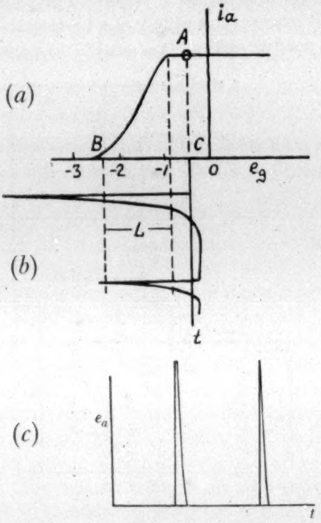


Figure 3.

This pulse-cleaning effect of the biasing is also accompanied by a pulse-shortening action corresponding to the pulse-shortening stage of a multistage amplifier with low input time constants. This is due to the fact that only the steepest parts of a negative grid pulse are amplified, and the low-frequency discharge of the grid (of time constant $R_g C_g$) is cut off.

There is further a most important pulse-equalizing effect. To discuss these effects we will consider two pulses which, as shown in figure 3 (*b*), are separated by an interval of time less than the recovery time $R_0 C_0$ of the counter. Because of this the second grid pulse is of smaller amplitude, but provided its peak extends beyond B in figure 3 (a) the output voltage as shown in figure 3 (c) will still be the same as for the first pulse. Further, the duration of the output pulses are much the same and become only slightly reduced when discharges fall close together.

The negative pulse impressed on the pentode grid consists of the negative rise (which, as we have seen, is of the order of 10^{-5} sec.) and of the return to zero potential, the time of which depends only on the grid time-constants. When $R_g = 5 \cdot 10^6$ Ω. and the grid capacity $C_g \cong 15\,\mu\mu$F., this return time may be taken as $7{\cdot}5 \cdot 10^{-5}$ sec., so that the duration of the whole pulse is a little less than 10^{-4} sec.

In order to examine the resolving-time for single counting with this amplifier we will refer to the diagram in figure 4. The curve R_0C_0 represents the recovery of the counter wire back to zero potential after a single count. As the time intervals between successive kicks decrease the input pulses decrease, and when they become less than BC, figure 3 (b), the amplitude of the output kicks becomes less than the normal equalized value. This is shown in figure 4, where the tips of the output pulses follow a curve such as f. A second pulse occurring after an interval less than t_l is not equalized, because the corresponding grid pulse is smaller than BC, but it may still be recorded. If the $\{i_a, e_g\}$ characteristic curve were of a rectangular shape, the equalizing action could be made perfect. If the horizontal line e represents the minimum size of output pulses which will be recorded, the intersection of this line with the line f determines the resolving-time τ_s of the amplifier without thyratron and recording circuit. τ_s

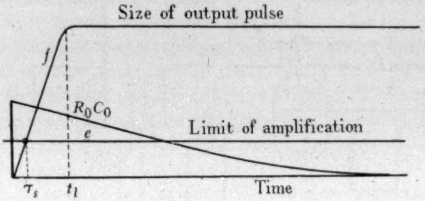


Figure 4.

It can therefore be seen that τ_s depends jointly on the value R_0C_0, on the pulse amplitude (which can be regulated by the adjustment of over-voltage) and on the amount of amplification. It may also be seen that as long as the pulses are single-peaked and of short duration their shape does not affect the resolving-power. The value of τ_s was measured by means of a pendulum breaking two contacts and found to be less than $\frac{1}{500}$ sec. for the single-valve amplifier and about $\frac{1}{1000}$ sec. for the 3-stage amplifier mentioned later, even when the value of R_0C_0 was as high as $\frac{1}{10}$ sec.

In order to deal with fast counting and smaller pulses a universal amplifier with two fore stages was built, making use of this special pentode characteristic. A description of this amplifier is given in the next section.

§ 4. CIRCUIT

As has already been pointed out, in order to obtain this characteristic, only the values of R_a and the actual h.t. voltage V_a are of importance. The simple circuit has been shown in figure 2. V_a, and consequently the saturation current, can be varied by means of a 500,000 Ω. potentiometer P_1. The V_a point must be decoupled by means of a large condenser. On increase of the saturation current the horizontal part of the characteristic is simply raised, the slope of the inclined portion remaining the same, and the knee moving nearer to zero potential. There will be an optimum point of adjustment according to the size of the pulse from the counter. It appears experimentally that this adjustment is not critical, however, and in the simplest case P_1 may be omitted and R_a may be connected directly to about 150 volts h.t. The potential of the screen grid is supplied from a voltage bridge across the source of h.t. and is kept steady by a 2μF. decoupling condenser. The potentiometer P_2 is essential for the adjustment of the working-point. The most sensitive part of the characteristic is the upper knee of the curve, and when the starting potential of the counter is to be determined the working point is shifted to this region, which is easily detected by a.-c. hum in the telephones.

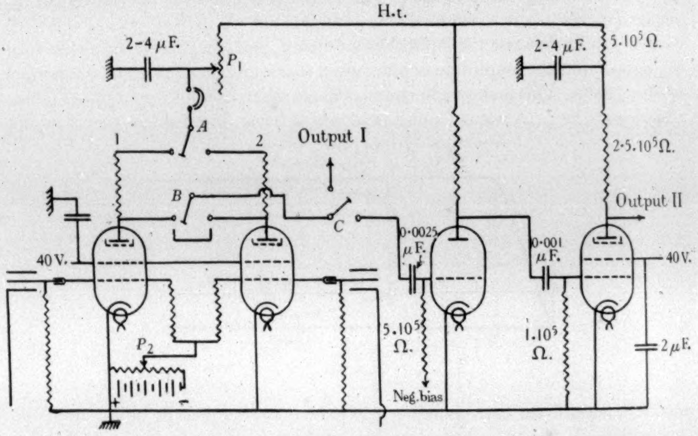


Figure 5.

As has been outlined in § 1, care must be observed in mounting the components of the grid input circuit. To reduce all stray capacities the leads should be short and it is desirable where possible to have the counter located very close to the grid terminal. Since in this circuit the bias suppresses small disturbances, shielding of the input leads may be omitted, the input capacity to earth being thus reduced.

To deal with very small input pulses and to allow of coincidence counting, a universal amplifier, figure 5, which requires only 4 valves, was built. The first two valves are SP4 pentodes arranged in parallel, each working exactly as in the simple circuit. The third valve serves purely as a voltage amplifier and may be any triode suitable for the purpose. An SPL with an amplification factor of 70 is satisfactory. The final valve is another SP4 pentode in a circuit differing from that for the input valves only by the omission of the two potentiometers P_1 and P_2. It is operated at zero grid bias. Switches *A* and *B* serve to adapt the amplifier for single counting by either of the input valves or for coincidence work. When the pulses are small switch *C* can be used to introduce the triode intermediate amplification and pentode final amplification.

§ 5. COINCIDENCE COUNTING

The adaptation of this circuit for coincidence incorporates the usual method developed by Rossi(4) and consists here of two equal SP4 valves. The working point *A* of each is slightly shifted into the saturation part of the characteristic.

As has been shown by Fussel and Johnson(5), pentodes are much superior to triodes for coincidence counting because the discrimination between partial and total coincidences is greatly improved. The form of the $\{i_a, e_g\}$ characteristic described in this paper seems to represent a further improvement. An amplifier using the normal pentode characteristic always shows small variations of output potential for partial coincidences. This effect is found to be reduced in the present amplifier for correct adjustment of the grid bias. The explanation is the fact that the pentode valve in the saturation region has such a low internal resistance that an anode pulse from the other valve is dropped to earth without affecting the output, while a coincidence input pulse of about 1 V. on the two grids causes a full anode swing.

To estimate the resolving-power of this amplifier for coincidences we note that pulses from the two counters will be separated if one pentode has returned to the saturation part, point *A* in figure 3, before the other starts to change. Now in § 3 we estimated that the duration of the whole change in the pentode was rather less than 10^{-4} sec., and we may take this figure as giving the order of magnitude of the resolving-time τ_c. A measurement of this quantity τ_c was carried out by use of the relation $\tau_c = D/2N_1N_2$, where D is the number of accidental coincidences between two counters when the number of discharges in each individually is N_1 and N_2. Values of τ_c from 5 to 7.10^{-5} sec. were obtained. It is of course obvious that the resolving-time depends on the grid time-constant of the pentode as determined by the appropriate resistances and capacities, but in addition it will be seen that the resolving-time will be improved if the input pulse is smaller so that the pentode remains a shorter time in the region of zero current. This result can be achieved by decreasing the over-voltages on the counters but this must not be carried too far, otherwise there will be a noticeable loss of coincidences due to a coincidence occurring while one counter has not yet recovered sufficiently from a previous discharge; see figure 4. Owing to this effect the resolving-power will depend on the number of single pulses in each counter.

τ_c
D
N_1, N_2

§ 6. PULSE-SHAPE

The shape of pulse developed may best be discussed by means of a simple equivalent circuit shown in figure 6. Distortion caused by operating across non-linear regions of the characteristic curve is not considered, because it does not seriously affect the final form of the pulses.

R_a, R_i, C_a The pentode is represented by switch *D*. R_a, R_i and C_a denote external resistance, internal resistance and plate capacity respectively. *C* is the coupling-
C_t, R_t condenser to the thyratron with grid capacity C_t and grid resistance R_t. As long as there is no pulse, switch *D* is connected to *a*. The plate is only slightly above earth-potential, because of the very low internal resistance R_i. The other side of the coupling condenser *C* is maintained at the grid-bias potential of the thyratron.

The steep wave-front of the discharge pulse is analogous to a sudden opening of switch *D* across the potentiometer. Plate-capacity C_a plus a small extra capacity (less than $C_t + 1/\omega^2 R_t^2 C_t$, where ω is the fundamental angular frequency of the first charge-rise) is quickly charged to the supply voltage, say 150 V. Taking the following approximate values—$C_a \cong 15 . 10^{-12}$ F., $C_t = 20 . 10^{-12}$ F., $R_g = 250{,}000\ \Omega$. and $\omega \cong 6 . 10^5$ (because the incoming pulse rises in a time $t_1 \cong 1 . 10^{-5}$ sec., § 2), we get for the second term of the above expression 2·6 cm. Hence the whole capacity, which is charged through R_a at the moment of the first rise, is of the order of $15 + 22 . 10^{-12}$
t_2 or $37 . 10^{-12}$ F. If $R_a = 250{,}000\ \Omega$. the time-constant t_2 is $37 . 10^{-12} \times 25 . 10^4$, i.e. about 10^{-5} sec. and is somewhat smaller than the time t_1 of the first pulse-rise. This
t_1 is of importance in this circuit, since if t_2 were greater than t_1 not all the pulses would rise to the full h.t. supply voltage of 150 V., and the limiting effect of the characteristic would become disturbed. The back pulse also, as will be seen later, would be eliminated. The conclusion from this consideration is that R_a should not exceed the value of 250,000 Ω.

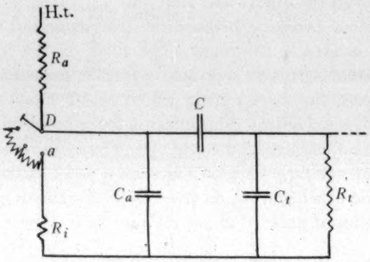


Figure 6.

This positive charging pulse, which is to set off the thyratron, is however reduced in size by the fact that an e.m.f. with resistance R_a is acting on an impedance, comprising C_a, C, C_t and R_t. This impedance is somewhat smaller than R_a. In a scale-of-two circuit R_t is still further reduced by the grid current of the tube when this is alight. Actually the amplitude of the pulse on the thyratron grid itself is reduced to about 30 or 40 V., a value which is quite sufficient, however, to overcome a normal bias of a few volts.

After passing the lower knee of the curve, figure 3, the system remains unperturbed for a very short time corresponding to the peak of the input pulse. In the equivalent circuit this is analogous to the time during which the switch is open. While the left side of C is still at high potential, the positively charged thyratron grid side starts to discharge through R_t with time-constant $R_t\,(C+C_t)$ and develops simultaneously a potential difference across C. This goes on until the working-point returns from the zero-current region to the active characteristic part with a time-constant determined by the input constants of the pentode grid. In the equivalent circuit this instant is represented by the connection of D to the highest-resistance point of the R_i potentiometer. The potential E_c that has developed across C now becomes divided proportionally between R_i and R_t and the grid of the thyratron must receive a negative potential given by $E_c\,R_t/(R_i+R_t)$. The thyratron grid, now negatively charged by the back pulse, discharges with a time constant $R_t\,(C+C_t)$. The pulse finally takes the form shown in figure 7 and represents a very favourable shape for triggering thyratron valves.

Figure 7.

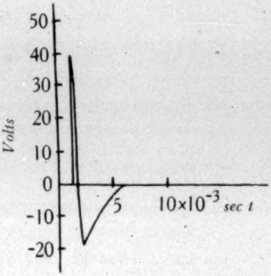


Figure 8.

In a recent paper$^{(6)}$ Morack has pointed out that this type of pulse is highly desirable. He was able to generate such a pulse by constructing a special thyratron input transformer with iron of several permeabilities. The back pulse helps to accelerate the de-ionization of the excited gas molecules after extinction of the arc. Lewis$^{(7)}$ has emphasized that a sharp, clean pulse can help in preventing the jamming effect. It was found experimentally that with the present amplifier this trouble is completely avoided, even at high rates of counting. It is also to be expected that this pulse-shape, when applied to a thyratron-couple with the improved anode-grid coupling circuit described by Lewis, may lead to a higher degree of resolution for high-speed counting than is at present available.

Oscillograph records were taken of the output pulses from the amplifier which would normally be fed to the thyratron grids. One such record of two close pulses is shown in figure 7. Owing to the rapid movement of the oscillograph spot the original photographs were not suitable for reproduction and the plate was obtained by careful retouching. A scale drawing of one of the pulses is given in figure 8. The actual rise and return to zero occurring within about 10^{-4} sec. would, if drawn correctly, be indistinguishable on the diagram, but they are shown separated for convenience. This superposition explains why this portion of the trace could just be seen on the oscillograph record whereas the remainder of the downstroke below the axis was invisible. It will be noticed that the back kicks of the two pulses shown in the plate are of different sizes; they were actually 19 V. and 15 V. It is clear from this that the two pulses must have occurred within the recovery time R_0C_0 of the counter, the actual kick from the counter being less for the second pulse, but yet the oscillograph record shows that the initial short positive pulse on the thyratron grids was in both cases the same, about 40 V. This is a good proof of the equalizing action of the amplifier.

§ 7. ACKNOWLEDGMENTS

In conclusion the author wishes to express his thanks to Lord Rutherford for permission to work in the Cavendish Laboratory and to Dr C. D. Ellis for his continued interest and encouragement. He is also greatly indebted to Mr W. J. Henderson for many fruitful discussions and help in the building and testing of the amplifier.

REFERENCES

(1) Johnson and Street. *J. Franklin Inst.* **215**, 239 (1933).

(2) Sven Werner. *Z. Phys.* **92**, 705 (1934), and **90**, 384 (1934).

(3) Pidgeon, H. A. "Theory of Multielectrode Tubes," *Elect. Engng*, Lond., p. 1488 (November 1934).

(4) Rossi. *Nature*, Lond., **125**, 636 (1930).

(5) Fussel and Johnson. *J. Franklin Inst.* **217**, 517 (1934).

(6) Morack, H. M. *Gen. elect. Rev.* **37**, 288 (1934).

(7) Lewis, W. B. *Proc. Camb. phil. Soc.* **30**, 543 (1934).

548.73:546.47-31

THE LATTICE-DIMENSIONS OF ZINC OXIDE

By C. W. BUNN

Received March 16, 1935. *Read in title June* 21, 1935

ABSTRACT. The lattice-dimensions of pure zinc oxide condensed from the smoke have been determined to a high degree of accuracy by means of X-ray powder photographs. Bradley and Jay's lattice-dimensions of quartz(4) and Siegbahn's X-ray wavelengths(5) being accepted as standards, the lattice-dimensions of zinc oxide are found to be:

$$\left.\begin{aligned} a_0 &= 3{\cdot}2426_5 \pm 0{\cdot}0001 \\ c_0 &= 5{\cdot}1948 \ \pm 0{\cdot}0003 \\ \text{Axial ratio } c_0/a_0 &= 1{\cdot}6020_0 \pm 0{\cdot}0001 \end{aligned}\right\} \text{at } 18^\circ \text{ C.}$$

Both the lattice-dimensions themselves and the axial ratio are lower than Finch and Wilman's electro-diffraction values(2) which are based on the X-ray lattice-dimension of gold, by amounts larger than the combined errors of both methods. There are thus real discrepancies, which are of the order of $\frac{1}{2}$ to 1 per cent.

In view of these discrepancies, Finch and Wilman's assumption that the apparent lattice-dimension of gold is the same for 40–60 kV. electrons as for X-rays may not be correct. The discrepancies therefore represent the combined possible discrepancies shown by gold and zinc oxide.

This does not apply to the axial ratio, which is independent of any reference substance. The difference between the axial ratios determined by electrons and by X-rays appears to point to a real difference in axial ratio between the surface and the interior of zinc oxide crystals.

§ 1. INTRODUCTION

Pure zinc oxide condensed from the smoke has been selected by Finch and Quarrell(1) as a reference substance in electron-diffraction determinations of crystal-lattice dimensions, and Finch and Wilman have recently determined its lattice-dimensions by reference to gold(2) using 40 to 60 kV. electrons. The values obtained in this way are considerably higher than those obtained previously by X-ray methods. The figures are collected in table 1, which includes a set of X-ray results that have appeared since Finch and Wilman's work.

Although the X-ray values differ considerably among themselves, they are consistently lower than the electron-diffraction values. Finch and Wilman conclude that the difference between the X-ray and electron-diffraction figures is due to a difference in the nature of the zinc oxide examined, and state that "this question could probably be settled, and the interests of accurate electron-diffraction analysis materially furthered, by an X-ray examination of pure condensed zinc-oxide smoke." This has now been done.

Table 1

Material	a_0 (A.)	C_0 (A.)	Axial ratio	Reference
X-ray method				
Zincite	3·22	5·20	1·608	W. L. Bragg(9)
Zincite	3·22	5·20	1·608	Aminoff(10)
Zincite	3·251	5·226	1·6077	Weber(11)
Zinc oxide (Merck)	3·242	5·176	1·596	Barth(12)
Zinc oxide (spectroscopically pure)	3·235	5·209	1·610	Fuller(13)
Zinc oxide	3·248	5·202	1·602	Ivannikov, Frost, Shapiro(14)
Electron-diffraction method				
Zinc-oxide smoke	3·258	5·239	1·607	Finch and Wilman(2)

§ 2. EXPERIMENTAL

Zinc-oxide smoke was made by the combustion of pure granulated zinc of forensic quality supplied by Prof. Finch. X-ray powder patterns were obtained by the Hull-Debye-Scherrer method, using a camera 9 cm. in diameter of the type described by Bradley and Jay(3). The camera was improved by the addition of brass screens which trap the primary beam soon after it has passed the specimen at the centre of the camera, the dimensions being such that none of the scattering which occurs inside the trap can possibly reach any part of the film; among other things, this has the effect of reducing the general background intensity, thus increasing the contrast of the lines and materially aiding the attainment of the highest accuracy of measurement.

In Bradley and Jay's type of camera, the exposed portion of the film is brought to an abrupt end by a knife edge, which thus registers a constant angle on every film. In the present instance this angle corresponds to a Bragg angle of 82°. For accurate determinations of lattice constants, only the reflections at large Bragg angles of 60 to 80°, which are well resolved into α_1 and α_2 doublets, are used. The distances between pairs of corresponding lines and the total length of the exposed portion of the film are measured, the angles being obtained by simple proportion. This eliminates errors due to shrinkage of the film.

The constant camera-angle is found by taking powder photographs of a substance whose lattice-dimensions are accurately known; Bradley and Jay have shown(4) that quartz is by far the best for this purpose. A sample of pure quartz, similar to that used by Bradley and Jay, was kindly supplied by Dr H. E. Buckley. Four powder photographs were taken with copper Kα radiation, four separately mounted specimens being used so that the corrections should be different. The
θ_k camera angle θ_k was found by calculating a value from each reflection in the range 60 to 80°, plotting these values against sin $2\theta/2\theta$, and extrapolating to sin $2\theta/2\theta = 0$. The mean value of θ_k from the four films was found to be $82{\cdot}125° \pm 0{\cdot}005°$. This error of ± 1 in 16,000 in the value of θ means that if the above value of θ_k is assumed, the lattice-dimensions of quartz can be obtained by the following method with an error of ± 1 in 60,000.

The lattice constants of a hexagonal crystal are found by first calculating an approximate value for the axial ratio c from several pairs of reflections by means of the equation

$$(d_{hkl})^2\,\{\tfrac{4}{3}\,(h^2+k^2+hk)+l^2/c^2\}=(d_{h_1k_1l_1})^2\,\{\tfrac{4}{3}\,(h_1{}^2+k_1{}^2+h_1k_1)+l_1{}^2/c^2\}.$$

This axial ratio is then used to calculate a_0 from all the reflections at large angles, thus a_0

$$a_0=d_{hkl}\sqrt{\left\{\frac{4}{3}\,(h^2+k^2+hk)+\frac{l^2}{c^2}\right\}}.$$

These values of a_0 are plotted against $\cos^2\theta$, and a linear extrapolation is made to $\cos^2\theta=0$. Bradley and Jay have shown that errors due to absorption of X-rays, thickness of specimen, and eccentricity of mounting are eliminated in this way. If there is any systematic one-sided deviation from the line on the part of values derived from reflections with large l indices, the axial ratio is in error, and another value must be tried.

In the present instance two powder photographs were taken with copper Kα radiation and two with cobalt Kα radiation, all with separately mounted specimens so that the corrections should be different. The zinc-oxide smoke gave very good photographs, with strong sharp lines.

There is no sign of line-broadening, and the crystals are therefore larger than 1000 A. in diameter. It so happens that with both radiations there is one strong doublet very near the end of the film at an angle of over 80°, and there are several more between 60° and 80°; this makes for great accuracy.

The films were measured on a Pye travelling microscope reading to 0·001 cm. Values of d/n were calculated from the Bragg equation $d/n=\lambda/2\sin\theta$ for each line. The pairs of values from α_1 and α_2 wave-lengths were averaged, the α_1 value being given twice the weight of the α_2 value on account of its greater intensity. A correction for refraction of X-rays was then made by means of Siegbahn's formula(5):

$$\text{True } d=d\,(\text{observed})\times\left(1+5{\cdot}4\,\frac{d^2\rho}{n^2}\,.\,10^{-6}\right),$$

where ρ is the density of the crystal (5·68 for ZnO). ρ

The indices of the reflections were found in the usual way, by means of a chart of the Hull and Davey type. The lattice constants were then found in the way already described. Details are given in tables 2 and 3. The value of the axial ratio c which best fitted all the films was found to be $1{\cdot}6020_0$. The error here is not likely to be greater than $\pm 0{\cdot}0001$.

The temperature near the camera was about 18° C. for all the films.

It may be noted that the final cobalt values are higher than the final copper values. The cobalt values could be reduced to the same level as the copper values by adopting a slightly higher axial ratio (1·6021), since the value from 213, which chiefly determines the final cobalt value, would be reduced by about the right amount; but a higher axial ratio would throw some of the other values too far out, and therefore has not been adopted.

Table 2. Powder photographs of ZnO taken with copper Kα radiation

$\alpha_1 = 1{\cdot}537395$ A. $\alpha_2 = 1{\cdot}541232$ A.

Re-flection	Intensity	Film I					Film II				
		θ (deg.)	$\frac{d}{n}$ (A.)	$\frac{d}{n}$ average	$\frac{d}{n}$ corrected	a_0	θ (deg.)	$\frac{d}{n}$ (A.)	$\frac{d}{n}$ average	$\frac{d}{n}$ corrected	a_0
213 α_1 α_2	Very strong	58·247 58·464	0·90401 0·90415	0·90405	0·90408	3·2396	58·226 58·463	0·90421 0·90416	0·90419	0·90422	3·2401
302 α_1 α_2	Strong	60·884 61·145	0·87988 0·87986	0·87987	0·87989	3·2400	60·880 61·117	0·87992 0·88009	0·87998	0·88000	3·2403
006 α_1	Weak	62·676	0·86524	0·86524	0·86526	3·2407	—	—	—	—	—
205 α_1 α_2	Strong	67·075 67·398	0·83462$_0$ 0·83472$_6$	0·83465$_7$	0·83467$_2$	3·2407$_0$	67·211 67·375	0·83485$_5$ 0·83486$_4$	0·83485$_9$	0·83487$_6$	3·2414$_9$
106 α_1 α_2	Medium weak	68·372 —	0·82691$_8$ —	0·82691$_8$ —	0·82693$_2$ —	3·2409$_8$ —	68·364 68·712	0·82696$_3$ 0·82707$_5$	0·82700$_1$	0·82701$_6$	3·2413$_1$
214 α_1 α_2	Medium	69·355 69·743	0·82145$_0$ 0·82142$_1$	0·82144$_0$	0·82145$_6$	3·2411$_4$	69·350 69·720	0·82147$_6$ 0·82154$_4$	0·82149$_9$	0·82151$_4$	3·2413$_7$
220 α_1 α_2	Strong	71·552 71·971	0·81034$_1$ 0·81040$_7$	0·81036$_4$	0·81037$_9$	3·2415$_2$	71·547 71·955	0·81036$_4$ 0·81048$_1$	0·81040$_3$	0·81041$_8$	3·2416$_7$
310 α_1 α_2	Strong	80·776 81·711	0·77877$_0$ 0·77875$_2$	0·77876$_6$	0·77877$_9$	3·2423$_1$	80·761 81·728	0·77880$_2$ 0·77871$_7$	0·77878$_0$	0·77879$_3$	3·2423$_7$
Extrapolated a_0						3·2426$_2$					3·2426$_0$

Table 3. Powder photographs of ZnO taken with cobalt Kα radiation

$\alpha_1 = 1·78529$ A. $\alpha_2 = 1·78919$ A.

Re-flection	Intensity	Film III					Film IV				
		θ (deg.)	$\frac{d}{n}$ (A.)	$\frac{d}{n}$ average	$\frac{d}{n}$ corrected	a_0	θ (deg.)	$\frac{d}{n}$ (A.)	$\frac{d}{n}$ average	$\frac{d}{n}$ corrected	a_0
211 α_1	Strong	59·269	1·03847	1·03852	1·03855	3·2384	59·227	1·03893	1·03892	1·03895	3·2396
α_2		59·468	1·03860				59·440	1·03890			
114 α_1	Medium strong	61·843	1·01246	1·01255	1·01258	3·2394	61·817	1·01271	1·01280	1·01283	3·2402
α_2		62·051	1·01271				62·023	1·01297			
212 α_1	Medium strong	65·409	0·98168$_2$	0·98171$_4$	0·98174$_3$	3·2400$_4$	65·375	0·98195$_0$	0·98199$_7$	0·98202$_6$	3·2409$_7$
α_2		65·672	0·98177$_5$				65·631	0·98209$_4$			
105 α_1	Strong	66·461	0·97366$_7$	0·97371$_3$	0·97374$_2$	3·2404$_6$	66·425	0·97393$_3$	0·97392$_6$	0·97395$_5$	3·2411$_7$
α_2		66·731	0·97380$_5$				66·716	0·97391$_5$			
204 α_1	Weak	69·544	0·95272	0·95272	0·95275	3·2404	69·496	0·95302	0·95302	0·95305	3·2415
300 α_1	Strong	72·564	0·93563$_6$	0·93569$_2$	0·93571$_9$	3·2414$_3$	72·535	0·93578$_5$	0·93581$_3$	0·93583$_9$	3·2418$_4$
α_2		72·934	0·93580$_2$				72·921	0·93586$_7$			
213 α_1	Very very strong	80·592	0·90481$_6$	0·90483$_4$	0·90485$_7$	3·2424$_0$	80·577	0·90485$_5$	0·90486$_0$	0·90488$_2$	3·2424$_8$
α_2		81·358	0·90487$_0$				81·358	0·90487$_0$			
	Extrapolated a_0					3·2427$_4$					3·2427$_2$

The discrepancy is probably due to refraction effects which cannot be precisely evaluated. Siegbahn's correction formula is for symmetrical reflection, i.e. the incoming and outgoing rays make equal angles with the crystal surface. This condition is usually not fulfilled on account of the development of a limited number of faces on the crystals; each line in a powder photograph is made up of many individual reflections which undergo varying deviations, the resultant of which may not correspond with that given by the formula (6). The last line on the cobalt films is from a totally different plane from that on the copper films, and the refraction corrections may not be quite right for either of them. However, since the extrapolated values of a_0 lie within the expected limits of error, this question need not be considered further; any inaccuracies due to these refraction effects are not likely to throw the final average value outside the limits given.

In the films taken with copper radiation, the last two reflections are from prism planes (310 and 220), and hence the extrapolated value of a_0 depends very little on the axial ratio adopted. These values have therefore been allowed double weight in taking a final average of a_0, which is found to be as follows:

$$a_0 = 3{\cdot}2426_5 \pm 0{\cdot}0001 \text{ A.}$$

Combining this with the axial ratio 1·6020 ± 0·0001, we get

$$c_0 = a_0 c = 5{\cdot}1947 \pm 0{\cdot}0004 \text{ A.}$$

In giving a possible error of ± 1 in 30,000 in the value of a_0, Bradley and Jay's lattice-dimensions of quartz are taken as standards. These in turn rest on Siegbahn's X-ray wave-lengths. Any error here will, of course, increase the absolute error of the present results. But for the present purpose of comparison with Finch and Wilman's electron-diffraction results, this does not enter into the problem, since those workers used as standard the lattice-dimension of gold, which rests on the same X-ray wave-lengths.

The c_0 dimension can be obtained with a little more precision by following the same procedure as for a_0, by means of the equation

$$c_0 = d\sqrt{\{\tfrac{4}{3}(h^2 + k^2 + hk)\, c^2 + l^2\}}.$$

The results are given in table 4.

Table 4. Powder photographs of ZnO. Determination of c_0

Reflection	c_0 (A.)		Reflection	c_0 (A.)	
	Film I	Film II		Film III	Film IV
213	5·1898	5·1906	—	—	—
302	5·1904	5·1910	211	5·1879	5·1899
006	5·1915	—	114	5·1895	5·1908
205	5·1916	$5{\cdot}1928_8$	212	$5{\cdot}1905_5$	$5{\cdot}1920_4$
106	$5{\cdot}1920_5$	$5{\cdot}1925_8$	105	$5{\cdot}1912_2$	$5{\cdot}1923_6$
214	$5{\cdot}1923_0$	$5{\cdot}1926_7$	204	$5{\cdot}1911_8$	$5{\cdot}1928_0$
220	$5{\cdot}1929_1$	$5{\cdot}1931_6$	300	$5{\cdot}1927_7$	$5{\cdot}1934_4$
310	$5{\cdot}1941_9$	$5{\cdot}1942_9$	213	$5{\cdot}1943_1$	$5{\cdot}1944_6$
	$5{\cdot}1946_5$	$5{\cdot}1946_8$		$5{\cdot}1949_2$	$5{\cdot}1948_6$

The results from films III and IV, taken with cobalt radiation, depend on the axial ratio to a smaller extent than the other results. Allowing them therefore double weight in taking a final average, we obtain

$$c_0 = 5{\cdot}1948 \pm 0{\cdot}0003 \text{ A.}$$

The limits of error for c_0 are wider apart than for a_0; this is because the lines nearest to the ends of the films are not basal-plane reflections.

§ 3. DISCUSSION

The lattice-dimensions found by the X-ray method differ from those found by the electron-diffraction method for similar material by amounts larger than the combined errors of both methods.

Table 5

	a_0	c_0	Axial ratio
X-ray method	$3{\cdot}2426_5 \pm 0{\cdot}0001$	$5{\cdot}1948 \pm 0{\cdot}0003$	$1{\cdot}6020_0 \pm 0{\cdot}0001$
Electron-diffraction method	$3{\cdot}258 \pm 0{\cdot}005$	$5{\cdot}239 \pm 0{\cdot}005$	$1{\cdot}607 \pm 0{\cdot}002$

There is thus a real difference between X-ray and electron-diffraction values; and not only are the lattice-dimensions themselves different, but also the axial ratios are different.

It must be remembered, however, that the electron-diffraction values for the lattice-dimensions rest on the assumption that the electron-diffraction lattice-dimension of gold is the same as the X-ray value. It now appears that this assumption may not be correct; if zinc oxide shows a discrepancy, gold may do the same. The discrepancies found here represent the combined differences of gold and zinc oxide. This does not apply to the axial ratio, which is independent of the reference substance.

The discrepancies recall the much larger discrepancies (up to 30 per cent) found by Davisson and Germer⁽⁷⁾ who used slow electrons with metallic crystals. On that occasion the apparent lattice-contraction was at first thought to indicate a real contraction on the crystal surface, which is revealed by electron waves on account of their very small penetration. But according to the calculations of Lennard-Jones and Dent⁽⁸⁾ the surface contraction on ionic crystals of the NaCl type is only of the order of 5 per cent, and is practically confined to the first layer of atoms; if the same applies to metal crystals, this explanation seems unlikely to be correct. It was considered that a more likely explanation of the apparent lattice-contraction was to be found in the refraction of electron waves by the crystal.

Here we have to account for apparent expansions of the order of $\frac{1}{2}$ to 1 per cent when fast (40 to 60-kV.) electrons are used. It is not very profitable to carry the discussion further at present, for there is insufficient evidence to point to a definite conclusion. Even the sign of the change in ZnO is uncertain, since the behaviour of gold is unknown. One fact, however—the difference between the axial ratios measured by the two methods—does appear to point to a difference in the crystalline material encountered by X-rays and by electrons. Since the penetrating power even of fast electrons is very much less than that of X-rays, this indicates a change of axial ratio at the surface. Any surface change of lattice dimensions would, of course, affect the *a* and *c* axes in different ratios.

§ 4. ACKNOWLEDGMENT

The author wishes to express his thanks to the Directors of Imperial Chemical Industries Limited for permission to publish this work, which was carried out in the Research Laboratories of their subsidiary company I.C.I. (Alkali) Limited.

REFERENCES

(1) Finch and Quarrell. *Proc. phys. Soc.* **46**, 148 (1934).
(2) Finch and Wilman. *J. chem. Soc.* p. 751 (1934).
(3) Bradley and Jay. *Proc. phys. Soc.* **44**, 563 (1932).
(4) Bradley and Jay. *Proc. phys. Soc.* **45**, 507 (1933).
(5) Siegbahn. *Spektroskopie der Röntgenstrahlen*, Eng. Trans. 1925, p. 26.
(6) Hägg and Phragmen. *Z. Krystallogr.* **86**, 306 (1933).
(7) Davisson and Germer. *Nature*, Lond., **119**, 558 (1927).
(8) Lennard-Jones and Dent. *Proc. roy. Soc.* **121**, 247 (1928).
(9) W. L. Bragg. *Phil. Mag.* **39**, 647 (1920).
(10) Aminoff. *Z. Krystallogr.* **56**, 495 (1921); **57**, 204 (1922).
(11) Weber. *Z. Krystallogr.* **57**, 398 (1922).
(12) Barth. *Norsk geol. Tidsskr.* **9**, 317 (1927).
(13) Fuller. *Science*, **70**, 196 (1929).
(14) Ivannikov, Frost and Shapiro. *C.R. Acad. Sci. U.S.S.R.* p. 124 (1933).

551.51.053.5

THE STRUCTURE OF THE IONOSPHERE

By Professor J. HOLLINGWORTH, M.A., D.Sc., M.I.E.E.

Received December 18, 1934. *Read in title June* 21, 1935

ABSTRACT. The following paper suggests a method of giving an approximate value to the intensity of ionization in the space between the *E* and *F* layers by measuring the relative group retardation of the two components of an echo from a pulse whose frequency is so close to the critical frequency of the *E* region that one component is reflected from the *E* layer and one from the *F*. It is pointed out that the value of the inter-layer ionization is of the utmost importance in determining the path of long-distance transmissions. From the results it is concluded that probably this inter-layer ionization is fairly uniformly distributed over the space between the layers, and is only a few per cent less in value than that at the top of the *E* layer.

§ 1. INTRODUCTION

As a result of the experiments of Appleton and others which have resulted in the magneto-ionic theory of wireless wave-propagation, it was originally held that the ionization of the atmosphere broadly consisted of two layers: the *E* or Kennelly-Heaviside layer at a height of about 100 km. above the surface of the earth, and the *F* or Appleton layer about 150 km. higher. The maximum ionization in the *E* layer being less than that in the *F*, signals for which the ionic content of *E* is not sufficient to return them break through abruptly to *F*, and this was verified by both the wave-length change and the pulse methods.

The actual value of the ionization in this intermediate region is, however, a decisive factor in the propagation of long-distance signals, as will be seen from the following argument. A signal of sufficiently high frequency to penetrate the *E* layer will, of course, be deflected by this layer; but if it emerges from it into an un-ionized region it will again revert to its original direction until it strikes the *F* layer, whereas if this intermediate space be ionized its direction at any point will be determined by the appropriate value of μ at that point.

The earlier workers on long-distance propagation adopted the original idea of an un-ionized space, since in their diagrams they show the signal path to the *F* layer as triangular, possibly with the apex slightly rounded off. On this supposition one would expect the principal ray to strike the earth at nearly grazing incidence; the exact angle being determined by a geometric triangulation between transmitter and receiver. Early in 1931, however, the author was engaged in measuring the angles of incidence of the signals from certain long-distance stations, and the results indicated that these angles were smaller, sometimes very much so, than this theory would lead one to expect. As a possible explanation it was suggested that the space between the *E* and *F* layers may be ionized to a degree only slightly less than that at the top of the *E* layer. Under these conditions the principal ray would be that which, subject to the necessary geometrical limitations, emerges from the top of the *E* layer at nearly grazing incidence, and consequently leaves the earth more steeply. If this were the case, the lower the frequency the more pronounced would be the effect, subject of course to the condition that the frequency was high enough to penetrate the *E* layer completely. On the waves of 15 to 20 metres used for commercial purposes the *E*-layer deflection could only be expected to be a few degrees, so that the effect would be much less marked and might easily be overlooked.

Recent measurements appear to have confirmed this supposition by showing angles of incidence of 75 to 85 degrees$^{(2,3)}$.

The same idea has since been put forward in various papers by other workers, chiefly Appleton and Ratcliffe, and is now generally accepted in principle. Their reasoning was based on the occasional appearance of echoes from a height lying between the accepted heights of the normal *E* and *F* layers; but from the nature of the method employed these echoes can be obtained only when the ionization of this region exceeds that of the top of the *E* layer. If such echoes are not present, all that can be said is that the ionization of this intermediate region is less than that of the top of the *E* layer, but no estimate can be given as to how much less. One investigator specifically refers to the "absence of the *e* region" on such occasions, and another shows a diagram with this portion of the curve relating ionization with height dotted. For long-distance propagation, however, it has been explained that it is most important to know something of the ionization of this region even when it is not capable of returning echoes, for the value of μ changes very rapidly for densities near the critical one, and in a recent paper by Appleton and Builder$^{(4)}$ an account is given of an effect observed occasionally, which seemed to the author to suggest a possibility of obtaining further information on this point.

In this paper Appleton and Builder show that when pulse observations at a definite frequency are being made on the *E* layer at a time when its ionization is steadily diminishing through the critical value for that frequency, reflections of the ordinary and extraordinary rays which have only just succeeded in penetrating the *E* layer may for a short time be actually reversed in time. They rightly attribute this to the reduction in group-velocity of the extraordinary ray, which can only just penetrate the *E* layer, relative to that of the ordinary ray whose critical density is much higher; and it occurred to the author that it might be possible to make an approximate computation of this time lag and see whether it could be accounted for by a discrete layer only 15–20 km. thick. The results appear to indicate very clearly that this is not possible; that in a layer whose ionization increases from zero at the bottom to the critical value some 15 km. higher up, and then falls off again, the distance over which there is appreciable reduction of group-velocity is so short that the time lag obtained is not comparable with that actually observed.

It must therefore be assumed that the ionization at the top of the *E* layer persists with only a slight diminution in value until the *F* layer is reached.

§ 2. ANALYTICAL METHOD

The standard formula for group-velocity U is U

$$\frac{1}{U}=\frac{1}{c}\frac{d}{dp}(\mu p),$$

c being the velocity of light *in vacuo*, p the periodicity, and μ the refractive index of the medium. c, p, μ

Hence if U and U' are the group-velocities of the ordinary and extraordinary rays, their coefficients of refraction being μ and μ' respectively, we have U' μ'

$$\frac{1}{U}=\frac{1}{c}\frac{d}{dp}(\mu p), \qquad \frac{1}{U'}=\frac{1}{c}\frac{d}{dp}(\mu' p).$$

Now if a velocity U operates over a distance ds, the time of passage is ds/U; i.e. $1/U$ represents the time occupied in travelling unit distance through a medium of constant electron-density of the appropriate value. The time lag between the two components thus becomes $1/U-1/U'$ per unit distance. ds

Therefore the differential time-lag rate between the ordinary and extraordinary rays is given by

$$\frac{1}{c}\left\{\frac{d}{dp}(\mu p)-\frac{d}{dp}(\mu' p)\right\}.$$

If $c=3.10^{10}$ this lag is in seconds per centimetre of path; but a more useful unit for this work is milliseconds per kilometre, in which case we get

Differential delay (msec./km.)

$$\frac{1}{300}\left\{\frac{d}{dp}(\mu p)-\frac{d}{dp}(\mu' p)\right\} \qquad \ldots\ldots(1).$$

A formula relating μ, p and N, the ionization-density, is given in the paper referred to, and is used here except that in view of recent mathematical investigations the expression $(-p^2/p_0{}^2-\frac{1}{3})$ for α has been replaced by $-p^2/p_0{}^2$. N

Subject to this condition the formula reduces to

$$\mu^2=1-\frac{2p_0{}^2}{2p^2-p^2p_T{}^2/(p^2-p_0{}^2)\mp\sqrt{\{p^4p_T{}^4/(p^2-p_0{}^2)^2+4p^2p_L{}^2\}}} \qquad \ldots\ldots(2),$$ p_T, p_L

where the symbols are as defined in the paper.(4)

There is no fundamental difficulty in evaluating the retardation formula from this expression, but the arithmetical work would be extremely heavy. Moreover the experimental measurement of the retardation times does not admit of a very high order of accuracy, so that we are concerned rather with the order of certain quantities than with their exact values. It is therefore permissible to use approximations throughout, and these have been made on the following basis.

If a rough calculation be made of the required retardation times, it will be found that the delay time when plotted against ionization-density remains quite small until the critical density is nearly reached, when it rises very rapidly. This is also evident from the curve in figure 1. The most important part of the curve is thus at densities differing from the critical density by first-order amounts, so that in the expression for $d(\mu^2)/dp$ we can after differentiation insert the critical relation for the extraordinary ray, namely

p_H

$$p'^2 - p_0^2 = pp_H.$$

When this has been done, the expression reduces to

$$\frac{d(\mu^2)}{dp} = \frac{2(2p' - p_H)}{p_0^2(1 + p_L^2/p_H^2)},$$

p' μ' and since $\mu = 0$ when p has its critical value p', we find that for any value of p close to p' we can write for the corresponding index of refraction μ'

$$\mu' = \left\{\frac{2(2p' - p_H)}{p_0^2(1 + p_L^2/p_H^2)}\right\}^{\frac{1}{2}} (p' - p)^{\frac{1}{2}} = \kappa(p' - p)^{\frac{1}{2}} \qquad \text{......(3)}.$$

Applying the standard group-velocity formula to this, we find that, since

$$\frac{d(\mu' p)}{dp} = \mu' + p\frac{d\mu'}{dp},$$

and since μ' is small and $d\mu'/dp$ large, the following result

$$\frac{1}{U'} = \frac{\kappa p'}{2c}(p' - p)^{-\frac{1}{2}}$$

is true for values of p close to p'.

For the purpose of evaluating the total delay through a layer of varying density we require the relation between $1/U$ and N, which can be obtained as follows. At the critical value for the extraordinary ray we have

$$p'^2 - p_0^2 = pp_H,$$

where

$$p_0^2 = 4\pi e N'/m.$$

By differentiating this with regard to p we obtain dN in terms of dp, i.e. we get $N' - N$ in terms of $p' - p$ where $N' - N$ is small; so that on substituting in equation (3) we finally obtain the following relation for the group-retardation rate for the extraordinary ray:

N'

A

$$\frac{1}{U'} = \frac{\sqrt{2p'}\,(2p' - p_H)}{cp_0(1 + p_L^2/p_H^2)^{\frac{1}{2}} A}(N' - N)^{-\frac{1}{2}} \text{ where } A^2 = \frac{4p_0^2}{N} = \frac{16\pi e^2}{m} \qquad \text{......(4)}.$$

Actually we require the difference between this and the rate for the ordinary ray; but as the latter is far from its critical frequency the approximate formula cannot be used. Another approximation is however possible here. In the original formula for μ it is found that when $p^2 > p_0^2$ the term $p^4 p_T^4/(p^2 - p_0^2)^2$ rapidly becomes small compared with $4p^2 p_L^2$. On this assumption the formula reduces to

$$\mu^2 = 1 - \frac{p_0^2}{p^2 + pp_L},$$

and by a process similar to the previous one we get for the delay rate of the ordinary ray

$$\frac{1}{U} = \left(2p - \frac{p_0^2 p_L}{p^2 + p_L^2}\right) \Big/ p.c.\left(1 + \frac{p_0^2}{p^2 + p_L^2}\right)^{\frac{1}{2}} \qquad \text{......(5)}.$$

There will, of course, be an intermediate stage in which neither of these approximations is valid. Here it is best to calculate a few points as accurately as possible and to use them to correct the other curves; but the whole of this range lies in a region where the delay is small, so that the possible error will not have any serious effect.

§ 3. NUMERICAL EVALUATION

In these calculations we shall assume that $p = 22 \times 10^6$ radians/sec., giving a wave-length of 85·7 metres which is of the same order as that used by Appleton and Builder. The critical density for the ordinary ray is then $1{\cdot}51 \times 10^5$, and for the extraordinary $0{\cdot}9467 \times 10^5$ electrons/cm.^{3}

Other constants for vertical transmission in the latitude of Slough are as follows:

$$p_H = 8{\cdot}302 \times 10^6, \qquad p_L = 7{\cdot}643 \times 10^6, \qquad p_T = 3{\cdot}24 \times 10^6.$$

Performing the necessary numerical work we finally arrive at the following results:

For values of N near to the critical value N′, for the extraordinary ray,

$$\mu' = 0{\cdot}00338\,(N' - N)^{\frac{1}{2}} \qquad \text{......(6)},$$

$$1/U' = 1{\cdot}383/(N' - N)^{-\frac{1}{2}}\ \text{(msec./km.)} \qquad \text{......(7)}.$$

For values of N far from N′,

$$\mu = 1 - 0{\cdot}488\,.\,N\,.\,10^{-5}\ \text{(ordinary ray)} \qquad \text{......(8)},$$

$$\mu' = 1 - 1{\cdot}007\,.\,N\,.\,10^{-5}\ \text{(extraordinary ray)} \qquad \text{......(9)},$$

$$\frac{1}{U} = \frac{44 \times 10^6 - 27{\cdot}8\,.\,N}{(48{\cdot}4 \times 10^{13} - 23{\cdot}2 \times N \times 10^8)^{\frac{1}{2}}} \cdot \frac{1}{2c}\ \text{(ordinary ray)} \qquad \text{......(10)},$$

$$\frac{1}{U'} = \frac{44 \times 10^6 + 118\,.\,N}{(48{\cdot}4 \times 10^{13} - 48{\cdot}7 \times N \times 10^8)^{\frac{1}{2}}} \cdot \frac{1}{2c}\ \text{(extraordinary ray)} \qquad \text{......(11)}.$$

The numerical results are given in the accompanying tables.

Table 1 (Ordinary ray)

$N \times 10^{-5}$	0·2	0·4	0·6	0·8	0·9	1·0
μ, equation (8)	0·95	0·9	0·84	0·78	0·75	0·72
μ, equation (2)	—	—	—	0·776	—	0·743
$1/U$	0·00347	0·0036	0·00383	0·00403	0·00413	0·00427

In order to check the accuracy of these approximate formulae the values of μ and $1/U$ have been tabulated so that the results given by the formulae overlap; and also a few values of μ calculated from the fundamental formula have been added.

Referring first to the ordinary ray, table 1, it will be seen that over the range considered the approximate and absolute values of μ agree so closely that there is little reason to doubt the accuracy of the $1/U$ formula; and in any case the delay on this ray is so small that slight errors have a negligible over-all result.

For the extraordinary ray, table 2, the calculation had to be made for a much larger number of values of N, owing to the approach to the critical point; but in the overlap period ($N=0{\cdot}8\times10^5$ and $0{\cdot}9\times10^5$) comparison with the accurate figures suggests that those results which are enclosed in brackets should be rejected.

As regards the relative values for $N=0{\cdot}94\times10^5$ it may be noted that it is quite likely that here the result from the approximate formula is more accurate than that from the general one, for when $N'-N$ is very small the expression for μ^2 in the general formula involves the small difference between two large numbers, and so requires an extremely large number of significant figures to yield high accuracy.

Table 2 (Extraordinary ray)

$N\times10^{-5}$	0·2	0·4	0·6	0·8	0·9	0·94
μ, equation (9)	0·89	0·77	0·62	[0·44]	[0·3]	—
μ, equation (2)	—	—	—	0·4	0·23	0·905
μ, equation (6)	—	—	—	0·433	0·23	0·874
$1/U$, equation (10)	0·00392	0·00477	0·0058	[0·0092]	0·0134	—
$1/U$, equation (7)	—	—	—	0·0014	0·0203	0·0503

The final results for $1/U-1/U'$ are plotted in figure 1, and give the differential group-retardation rate between the rays for the given range of densities. Now the exact distribution of density with height in the layer is not known, and it will be assumed that the density is proportional to the height above the point at which ionization commences, since this will at any rate give an approximation to the real

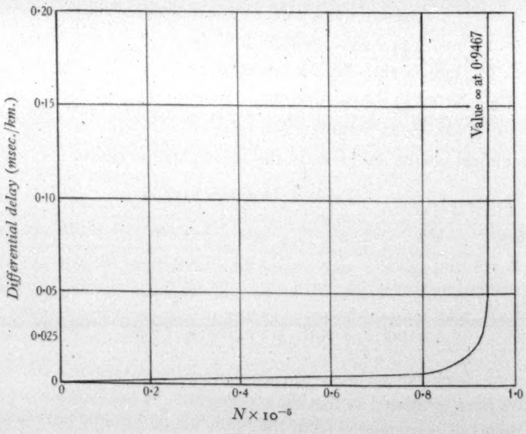


Figure 1.

figures. Below this height, of course, the differential retardation is always zero. Now if we have a layer graded in this way with a maximum density N', the total delay will be twice the integral of this curve up to the point $N=N'$ multiplied by the thickness of the layer, the factor 2 arising from the fact that the ray passes twice through the layer. This integration can be performed graphically to within a short distance of the critical density, where the curve becomes too steep for accurate evaluation.

It is, however, important to note that although the delay rate is infinite at the critical density yet in a layer graded up to this density the total delay time is finite, since it is an integral of the form $\int \frac{dN}{(N-N')^{\frac{1}{2}}}$, so that the total delay time can be calculated right up to this value, and the results of this calculation are given in the third column of table 3. For convenience the densities are given not in their absolute values but as percentage departures from the critical density, $N=0{\cdot}9467\times10^5$, since in this form they are probably roughly applicable to other critical densities of the same order. An inspection of the traces given by Appleton and Builder will show that the delays under consideration are of the order of 1 msec.

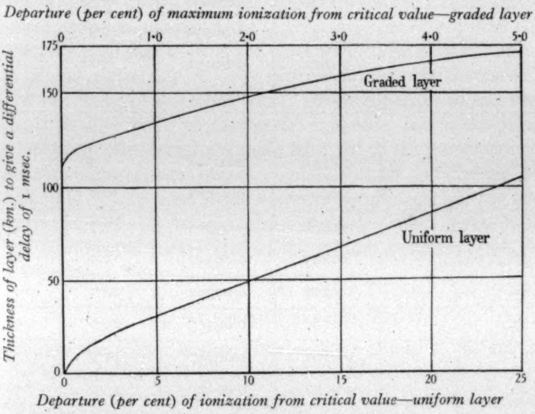


Figure 2.

To get the above results in the most suitable form, the figures in table 3 have been used to determine the thickness of layer required, both uniform and graded, to give a delay of 1 msec., and the figures obtained have been plotted in figure 2.

On examination of these results it will be seen that to obtain the necessary delay with a graded layer, even if the maximum value is *exactly* the critical value, would require a thickness of 106 km. Now from the fact that signals returned from the E layer show only small differences in their delay times over a considerable range of frequencies, it is clear that the graded part of this layer must be comparatively thin. Consequently to obtain a delay of the observed value the ray must pass through a much greater distance of ionized air.

Table 3*

Density (per cent departure from critical value)	Delay (msec./km.)		Thickness giving delay of 1 msec. (km.)	
	Uniform layer	Graded layer	Uniform layer	Graded layer
0	∞	0·00471	0	106·0
0·01 of 1 per cent	0·43	0·0045	1·16	112·0
0·02 ,,	0·308	0·0041	1·63	113·0
0·04 ,,	0·213	0·00436	2·35	114·0
0·07 ,,	0·161	0·00431	3·10	116·0
0·12 ,,	0·122	0·00424	4·10	117·0
0·17 ,,	0·10	0·00418	5·0	120·0
0·50 ,,	0·0591	0·00394	8·46	126·0
0·71 ,,	0·049	0·00383	10·0	130·0
1·23 ,,	0·0366	0·00363	13·7	137·0
1·76 ,,	0·0297	0·00347	17·0	145·0
2·80 ,,	0·0227	0·00322	22·0	155·0
3·87 ,,	0·0187	0·00303	27·0	165·0
4·93 ,,	0·0162	0·00287	31·0	170·0
10·20 ,,	0·0102	0·00227	49·0	220·0
15·40 ,,	0·007	0·00178	71·5	280·0
20·80 ,,	0·006	0·00157	83·0	318·0
26·0 ,,	0·0048	0·00125	109·0	400·0

* All the above results are based on a periodicity of 22×10^6 and a critical frequency for the extraordinary ray of $0{\cdot}9467 \times 10^5$. Probably however they are approximately true over an appreciable range of values.

Moreover the curves show that the delay decreases very rapidly for small departures from the critical density and then changes much more slowly, so that the effect could not be obtained from small patches of layer nearly at the critical density unless we are prepared to admit that the ionization is sufficiently steady for such small patches to retain their ionization for considerable periods of time at a value not differing from the critical value by more than very small fractions of 1 per cent.

The natural assumption appears to be that the space between the E and F layers is ionized fairly uniformly to a value not much less than that at the top of the E layer, or alternatively that the ionization of the layer as a whole rises very rapidly at first for a depth of from 10 to 20 km. and thereafter increases at a much slower rate. The observed delays could then be obtained whenever the density of the layer lies within from 10 to 15 per cent of the critical value, which seems a much more reasonable assumption.

It may be argued that the phenomenon referred to by Appleton and Builder appears to be only rarely observed, and so may be due to an unusual and abnormal state of the layer. But in many of the routine traces taken at the Radio Research Station at Slough it will be noticed that immediately after the break-through to the F layer the trace is convex to the origin for a short time, showing an abnormal delay, which is clearly due to the same cause, around the critical frequency. The values of these delays as shown on the traces are generally less than 1 msec.; but owing to the very rapid variation of the delay curve at very small departures from the critical density, extremely accurate instrumental correlation would be needed to obtain the full range of this delay bend. In the apparatus at present in use the wave-length is steadily changing during the emissions, and as the pulses are sent out at the rate of 50 per second not more than one at the utmost would be likely to lie within this highly critical period. For this reason also it seems unprofitable to attempt to build up any exact form of layer-structure from existing observations, though this could probably be done from the calculations given in the present paper if the instrumental survey around the critical frequency could be made of a more rigid nature.

§4. ACKNOWLEDGMENTS

In conclusion the author wishes to express his thanks to the Radio Research Board of the Department of Scientific and Industrial Research for permission to make use of some of the results obtained at the Radio Research Station, Slough; and also to Dr Mary Slow and Mr Naismith for their help and suggestions.

REFERENCES

(1) Hollingworth. *J. Instn elect. Engrs*, **72**, No. 435 (March 1933).
(2) Walmsley. *J. Instn elect. Engrs*, **74**, 543 (1934).
(3) Wilkins. *J. Instn elect. Engrs*, **74**, 582 (1934).
(4) Appleton and Builder. *Proc. phys. Soc.* **45**, Part 2, No. 247 (1933).

537·322.2 : 546.74

THE LONGITUDINAL THERMOELECTRIC EFFECT: (II) NICKEL IN LONGITUDINAL MAGNETIC FIELDS

BY T. H. PI, M.S. (Yenching) AND WILLIAM BAND, M.Sc.,
Professor of Physics, Yenching University, Peiping, China

Received October 1, 1934. *Read in title May* 3, 1935

ABSTRACT. The e.m.f. produced by asymmetrical temperature-distributions in a pure nickel wire is examined and found to agree well with the formula

$$\mathbf{F} = t\mathbf{G} + a\,(\mathbf{G}.\mathbf{G})^{\frac{1}{2}}\,\mathbf{G} + b\,(\mathbf{G}.\mathbf{G})\,\mathbf{G},$$

where **F** is the potential-gradient at the point where **G** is the temperature-gradient. The two constants have the values

$$10^9.a = -1{\cdot}80 \text{ V.-cm.}/(^\circ \text{ C.})^2,$$

$$10^{11}.b = 1{\cdot}85 \text{ V.-cm.}^2/(^\circ \text{ C.})^3,$$

for zero magnetic field. For small fields both constants are reduced in absolute value, reversing sign at 40 gauss, reaching a maximum of this opposite sign at 200 gauss, and thereafter returning asymptotically towards zero. The connexion of these results with thermomagnetic phenomena is discussed.

---

§ 1. INTRODUCTION

E

ACCORDING to the results published in the first paper of this series, on copper[1], the thermoelectric e.m.f. E in a homogeneous wire is expressible simply as a function of the temperature-distribution:

a, *b*

$$E = a\Sigma\,(G.dT) + b\Sigma\,(G^2.dT),$$

G, *T*

where G is the temperature-gradient and T the temperature, and where in the first term on the right-hand side the positive magnitude of G must be used without regard to the sign of the gradient. The potential-gradient thus contains two terms, a quadratic and cubic in the temperature-gradient, in addition to the linear Thomson term.

In the present work we find that the thermoelectric e.m.f. in pure nickel can be represented by the same function with different values of the constants a and b, and that these constants become functions of the magnetizing field when the whole temperature-distribution is placed in a uniform field.

The e.m.f. for copper is opposite in direction to that for nickel, in zero magnetic field, and the values for nickel are much greater than those for copper.

§ 2. APPARATUS AND TECHNIQUE

The specimen was pure thermo-element nickel wire of diameter 0·5 mm. It was stretched by a weight of 500 gm. at each end of the temperature-control system, the specimen making a right-angle bend in two clamps before being taken to the bath, which comprised insulating oil kept at constant uniform temperature.

The temperature-control system consisted of a brass rod with a central bore, at both ends of which were water cooling-jackets at the same cool temperature. The heater was situated close to the inner end of one of the cooling-jackets; it was a nichrome wire wound non-inductively on a truncated slate pyramid surrounding the brass rod coaxially.

This system was adapted to slide into the core of a water-cooled solenoid of the Moullevigen form to give as nearly uniform a field as possible throughout the length of the temperature-distribution.

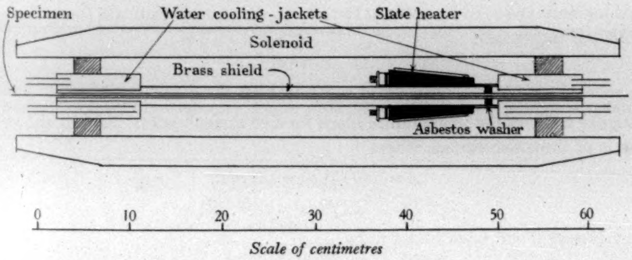


Figure 1. Diagram of the apparatus, showing the temperature-control system in position along the axis of the solenoid.

Temperatures were recorded entirely by movable thermocouples sliding in the bore of the tube. The same technique as that described in the previous paper was employed, and the same remarks concerning precision also apply.

The e.m.f. was measured on a specially adapted potentiometer of thermocouple type manufactured by the Cambridge Instrument Co. Ltd. As supplied this instrument gives readings only to 10μV., and is not sensitive enough for our purpose. A shunt and series resistance were added, as in the *K*-type potentiometer, to increase sensitivity. The resistances were non-inductively wound double-silk-covered Eureka wire no. 28 (for the shunt) and no. 40* (for the series resistance); the resistance of the shunt was made exactly one-nineteenth of the total potentiometer resistance, and that of the series resistance was adjusted so that when both were present in the circuit there was no change in the working current. These adjustments were made by means of a *K*-type potentiometer with very good precision.

---

* British standard gauge.

A single-pole double-throw switch enabled us to use the potentiometer either in its original form or in its new form; it gave excellently consistent results both ways, and in the sensitive position estimates down to $0{\cdot}1\,\mu$V. were easily and reliably made. The scale reading was $0{\cdot}5\,\mu$V. per division, which is even better than that given by the *K*-type potentiometer. The potentiometer was standardized against a Weston standard cell in a thermostat. A Cambridge high-sensitivity suspended-coil galvanometer was needed to indicate the potentiometer balance, the galvanometer ordinarily used not being sensitive enough to match the increased sensitivity of the potentiometer scale.

With this high sensitivity great care had to be taken with all binding posts, switches, etc., and all electrical connections were separately shielded in grounded tubes. The ice available was not pure enough to provide perfectly constant temperatures, and the ends of the specimen were therefore taken to a bath of good insulating oil where they were connected to the potentiometer leads. The oil was well stirred, and the bath was surrounded by a good heat-insulating case. The junctions were about 2 cm. apart in the oil; no corrosive action was found between the oil and the metals.

§ 3. RESULTS

Figure 2 shows the temperature maps for five typical cases. Table 1 shows the analysis of these five curves, where

A $$A = \Sigma\,(dT)^2/dx = \Sigma G\,.\,dT,$$

B $$B = \Sigma\,(dT)^3/dx^2 = \Sigma G^2\,.\,dT.$$

The values of *a* and *b* are found to be

$$a = -1{\cdot}80\,.\,10^{-9}\ \text{V.-cm.}/(^\circ\,\text{C.})^2,$$

$$b = \ \ 1{\cdot}85\,.\,10^{-11}\ \text{V.-cm.}^2/(^\circ\,\text{C.})^3,$$

and the values of $(aA+bB)$ in the table are the theoretical values of the e.m.f. calculated from these values of *a* and *b*. The last column shows the measured values of the e.m.f. Positive e.m.f. is directed from the steep gradient to the small gradient in the specimen, which is the direction of the e.m.f. in copper.

Table 1

Curve	$A\,.\,10^{-2}$	$B\,.\,10^{-4}$	$aA+bB$ (μV.)	E (μV.)
1	24·20	7·75	−2·9	−3·0
2	39·37	18·95	−3·6	−3·5
3	41·71	19·94	−3·8	−3·9
4	44·55	19·50	−4·4	−4·5
5	69·93	43·95	−5·2	−5·2

The effect produced by changes in tension has been observed but not examined in detail.

The effect produced by magnetization is considerable. Figure 3 shows e.m.f. plotted against magnetizing field for the five temperature-distributions of figure 2.

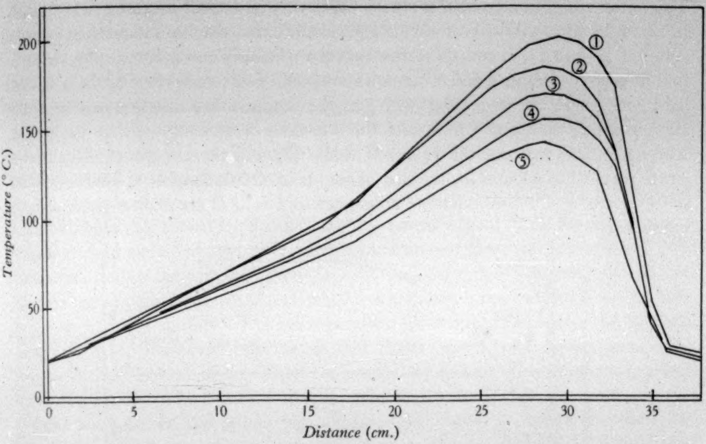


Figure 2. Typical temperature maps, the distributions of temperature are numbered to correspond with the values of e.m.f. recorded in table 1 and figure 3. The curves are drawn as split up for analysis into rectilinear sections.

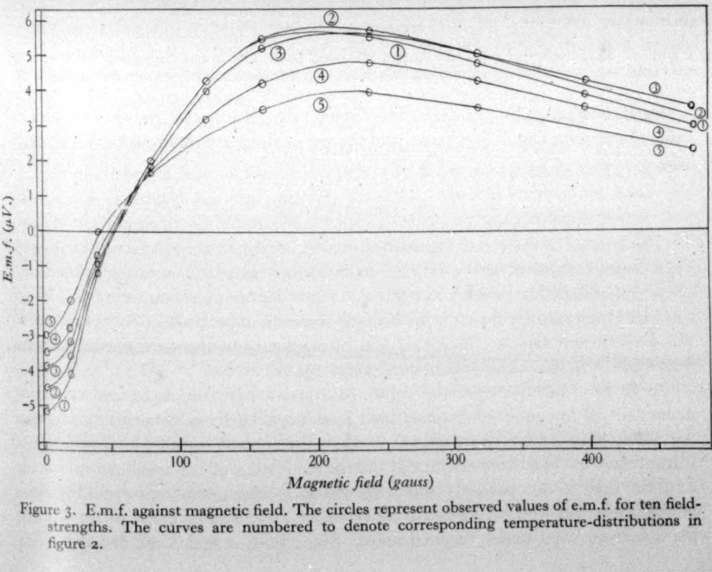


Figure 3. E.m.f. against magnetic field. The circles represent observed values of e.m.f. for ten field-strengths. The curves are numbered to denote corresponding temperature-distributions in figure 2.

The initial e.m.f. in zero field is always negative, but a small magnetic field quickly reverses the sign. At fields around 200 gauss the e.m.f. reaches a maximum positive value of around 5 μV. and thereafter decreases linearly towards zero. We wish to find how the constants a and b vary with magnetic field; each value of the magnetic field was therefore associated with all the temperature-distributions and the corresponding values of e.m.f., and the constants were computed for each field. The agreement between the values of $(aA+bB)$ and the experimental values of e.m.f. was just as good as for the case of zero field shown in table 1. The results for the values of a and b as functions of magnetizing field H are shown graphically in figure 4.

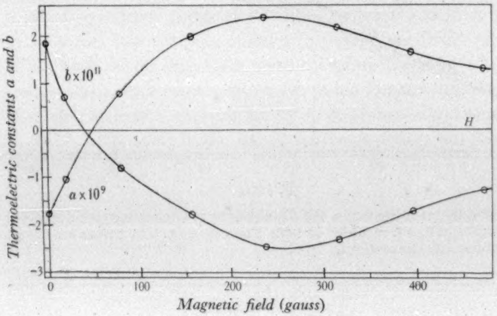


Figure 4. Thermoelectric constants against magnetic field. The circles represent the results of calculations for all temperature maps obtained at the various values of magnetizing field.

Each constant undergoes a reversal of sign at almost 40 gauss, rises to a maximum value of opposite sign at 200 gauss, and thereafter decreases asymptotically towards zero.

§ 4. DISCUSSION

The success of the foregoing analysis makes it evident that there can have been no recrystallization of the specimen to introduce spurious non-homogeneity at uncontrollable points in it. A test was also made for recrystallization by displacing the wire longitudinally through the heating-system; no permanent change of e.m.f. was caused, and this fact indicated real homogeneity of the wire throughout the temperature-ranges used, which were under 200° C.

As in the case of copper the values of a and b have been assumed to be independent of temperature because they have been taken as constant throughout the whole temperature-distribution. In the present work another assumption not immediately obvious has been made: the magnetization of the specimen under the uniform field is not perfectly uniform, owing to changes of susceptibility with temperature, but we have assumed the values of a and b to be constant throughout the specimen even under magnetization. Since both a and b are found to vary with magnetization this assumption cannot be exactly true. On the other hand the success of the analysis indicates the assumption to be at least approximately true. Further, the greatest temperature-differences existing in the wire are less than 200° C., and nowhere is the temperature near the Curie point for nickel; the susceptibility varies most at the maximum temperature, where the temperature curves are nearly symmetrical and temperature-variations contribute least to the e.m.f. This however is a difficulty which any work of higher precision than the present would have to overcome.

The curves of figure 3 should be compared with the curves of thermomagnetic e.m.f. obtained for nickel by Broili(2), and by Tao and Band(3). In Broili's work the magnetic field extended over only one temperature-gradient, temperature was kept constant through the varying magnetic field, and the return gradient was well out of the magnetic field. Broili stated that his e.m.f. was independent of the gradient but was a function only of the temperature-difference existing in the uniform magnetic field. This evidently means that Broili's e.m.f. was caused entirely by magnetic modification of the Thomson coefficient t. This Thomson e.m.f. would mask any smaller effect due to changes in a and b in Broili's apparatus. In our present work, since the whole temperature-distribution is within the field, the Thomson coefficient, even though changed by the magnetic field, contributes nothing since the end temperatures are equal.

This last statement is again subject to the following qualification: owing to temperature differences the magnetization of the specimen is not uniform, and hence the value of t is variable in the specimen even for a uniform magnetizing field. In our case the Thomson e.m.f. will still integrate to zero, but it is fruitful to discuss whether differences in magnetization of this kind are responsible for thermomagnetic phenomena.

Thus in the work of Chang and Band(4) it was suggested that the thermomagnetic e.m.f. was directed from parts of the specimen (in that case iron) with higher spontaneous magnetization to parts with lower spontaneous magnetization. The e.m.f. in nickel has the opposite direction. Now if there is an e.m.f. due merely to differences in longitudinal magnetization, we must remember that there are the same differences of magnetization running in reverse directions as the wire emerges from the magnetic field at either end (in the one case at a low temperature, and in the other at a high temperature, if we consider Broili's arrangement). These differences are much greater than the small differences due to the dependence of susceptibility on temperature within the magnetic field; in fact the effect of temperature is merely to destroy the symmetry of the magnetization-distribution. In the arrangement of the present work there is also this asymmetry of magnetization-distribution, but we can hardly consider the idea analytically.

Evidently thermomagnetic e.m.f. is more simply explained by magnetic modification of the three thermoelectric constants; the influence of differences of magnetization due to changes of susceptibility with temperature is probably confined to the smaller hysteresis phenomena discussed in the work of Chang and Band already referred to above.

Finally we point out that our magnetic fields are not strong enough to eliminate the effect of ferromagnetization as distinct from spontaneous or true magnetization; probably the changes in *a* and *b* shown in figure 4 are due entirely to ferromagnetic changes or reversals of the elementary magnetic axes of crystals. Whether there will remain some effect on *a* and *b* after saturation of the reversal process cannot be determined without fields of the order of 3000 gauss, as used by Englert in his study of resistance-changes in nickel$^{(5)}$.

REFERENCES

(1) FENG, P. C. and BAND, W. The longitudinal thermoelectric effect: (I) copper. *Proc. phys. Soc.* **46**, 515–22 (1934).
(2) BROILI, HEINZ. Ferromagnetismus und elektrische Eigenschaften: (V). *Ann. Phys., Lpz.*, **14**, 259–72 (1932).
(3) TAO, S. C. and BAND, W. Some thermomagnetic effects in nickel and iron. *Proc. phys. Soc.* **44**, 166–8 (1932).
(4) CHANG, W. Y. and BAND, W. Thermomagnetic hysteresis in steel. *Proc. phys. Soc.* **45**, 602–9 (1933).
(5) ENGLERT, ERNO. Ferromagnetismus und elektrische Eigenschaften: (VI). *Ann. Phys., Lpz.*, **14**, 589–612 (1932).

537.322.2:546.621

THE LONGITUDINAL THERMOELECTRIC EFFECT: (3) ALUMINIUM

By M. K. LI, M.S. (Yenching) and WILLIAM BAND, M.Sc.,
Professor of Physics, Yenching University, Peiping, China

Received October 6, 1934. *Read in title May* 3, 1935

ABSTRACT. The thermoelectric e.m.f. produced in aluminium wires is measured by a Paschen galvanometer. The temperature-distribution is analyzed by means of the formula previously used, but without success; alternative formulae also fail. The e.m.fs. observed were less than 0·3μV.; they were in the same direction as those previously found in copper wire.

§ 1. INTRODUCTION

The e.m.f. in commercially pure aluminium wire produced by asymmetrical heating to temperatures always under 180° C. has been observed and measured. The temperature-controlling system is the same as that used in the work of Feng and Band⁽¹⁾ on copper, but the e.m.f. was so minute that a Paschen galvanometer (made by the Cambridge Instrument Co. Ltd.) had to be employed to detect it. The e.m.f. was in the same direction as that for copper, directed from the steep gradient to the small gradient within the specimen; it was never more than 0·3μV.

The Paschen galvanometer was calibrated by means of the modified thermocouple potentiometer as described in the paper by Pi and Band⁽²⁾, which gave results that could be read to 0·5μV., and estimated to 0·1μV.* The deflection sensitivity of the galvanometer was adjusted and maintained at 0·0032μV./mm. at the scale-distance of 1 m.

Even with this high sensitivity, the e.m.f. values obtained are satisfactorily steady during the mapping of the temperature-distribution by means of the thermocouple probe⁽¹⁾; the mean deviation of the e.m.f. value during probing was usually between 1 and 2 per cent.

The aluminium specimen was connected at its ends to two thick copper leads, the connexions being made in a temperature bath containing good machine oil; it is more important that the junctions should be at identical temperature than that they should be at some definitely known temperature. Leakage and corrosion, and impurities, in an ice bath caused quite uncontrollable irregularities in the e.m.f.

The thick copper leads passed through grounded shields and were connected directly to the galvanometer. A switch would make the circuit impracticably complicated, owing to contact charges. The resistance of the specimen being small compared with that of the galvanometer, the latter is well short-circuited; it was found that several hours were needed for the galvanometer reading to return to its open-circuit value after the circuit had been completed. Throughout the work this circuit was left undisturbed; the zero reading was determined before the investigation of each temperature map by leaving the whole system at room-temperature overnight, and was found to remain excellently constant for the whole period of several months during which it was used.

---

* The potentiometer measured the standard voltage across a standard resistance (made by the Hartmann Braun Co.) and a tapping on this was applied to the galvanometer.

§ 2. RESULTS

Eight or ten different temperature-distributions were examined for each of the four different tensions due to loads of 0, 1, 2 and 3 kg. on the ends of the wire, which had a diameter of 1·219 mm.

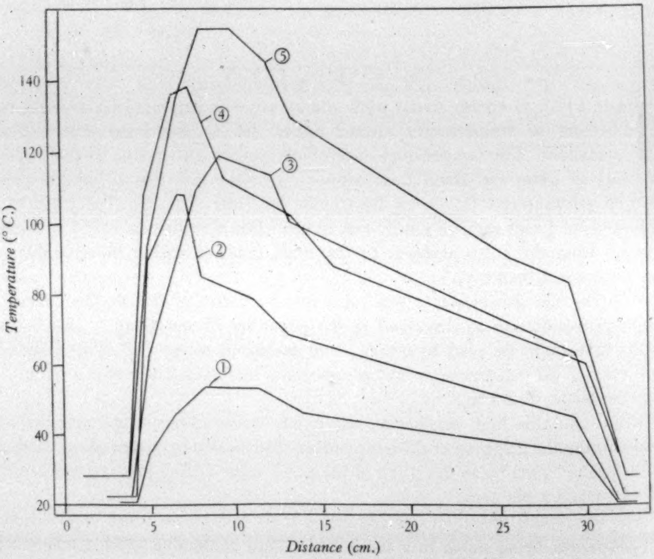


Figure 1. Temperature maps: wire under zero tension. The thermo-e.m.fs. (μV.) were as follows: (1) 0·0320; (2) 0·0454; (3) 0·2426; (4) 0·0662; (5) 0·2525.

Typical temperature maps are shown in the figure. These were analysed by the same method as that used for copper, but absolutely no correlation could be found between the observed e.m.f. values and the integrals from the temperature maps. It was then thought that if the constants a and b in the formula(1)

$$E = a \int G \, . \, dT + b \int G^2 \, . \, dT$$

were allowed to vary with temperature along the wire, a better correlation might be achieved.

Formulae of the form

$$E=a'\int T.G.dT+b'\int T.G^2.dT$$

and also

$$E=a'\int T.G.dT+b'\int T^2.G^2.dT$$

were used, but with little better success. Not more than three of the maps under any one tension-state could be correlated by any one choice of constants a' and b'.

We conclude that the asymmetrical temperature-distribution is not the controlling factor in the production of the longitudinal thermoelectric e.m.f. in the case of our specimen of aluminium.

There are several possible explanations for the above results. First, what was an apparently uniform and homogeneous wire may have had some impurity non-uniformly distributed in its material; in this case a sudden change in constitution at one point within the high-temperature part would give rise to an e.m.f. independent of the temperature except at that point. Secondly, as a result of previous heating to a high temperature in one part of the wire, the aluminium may have become recrystallized, a discontinuity of structure being thus introduced at one or more points along the specimen. From previous work(3, 4, 5) it is known that aluminium has recrystallized after being heated for several hours at 500° C., and has done so rapidly at 600° C. In our present work no part of the wire was ever heated above 180° C., and even this temperature was not maintained for more than a few hours at a stretch. It does not seem possible that recrystallization could have taken place, unless at some time in its previous history the wire had been heated to a high temperature in disconnected parts—a highly improbable supposition.

The only fruitful suggestion we can make is that there is an allotropic form of aluminium with a transition temperature somewhere below 180° C., accompanying a structural change so slight that the thermoelectric powers of the two forms differ by not more than about 0·005 μV. per degree of difference in temperature. One of us (W. B.) has succeeded in obtaining a correlation of the present results on the basis of this assumption, but a direct experimental test of the hypothesis will be made before the results of this analysis are published.

REFERENCES

(1) Feng, P. C. and Band, W. The longitudinal thermoelectric effect: (1) copper. *Proc. phys. Soc.* **46**, 515 (1934).

(2) Pi, T. H. and Band, W. The longitudinal thermoelectric effect: (2) nickel, p. 852 of this volume.

(3) Burgers. Recrystallization of aluminium. *Z. Phys.* **59**, 9–10, 651–5 (1930); **67**, 605–78 (1931).

(4) Schmid and Wissermann. Recrystallization of aluminium. *Metallwirtschaft*, **10**, 409–10 (1931).

(5) Karnop and Sarks. Kinetics of recrystallization in copper and aluminium. *Z. Phys.* **60**, 464–80 (1930).

537·322.2:546.621

THE LONGITUDINAL THERMOELECTRIC EFFECT: (4) A FURTHER STUDY OF ALUMINIUM

By WILLIAM BAND, M.Sc., Professor of Physics,
Yenching University, Peiping, China

Received October 12, 1934. *Read in title May* 3, 1935

ABSTRACT. Results obtained in work by Li and Band$^{(1)}$ are analysed by means of the formula$^{(2)}$

$$E = a\int G.dT + b\int G^2.dT + e.DT,$$

where e is the excess in thermoelectric power of normal aluminium over that of a postulated allotropic form having a transition temperature T_c below 180° C. with delayed reverse transition; DT is the temperature difference between the junctions of the two forms of metal. Excellent agreement for all tensions is found. Under zero tension the values found are

$$a = +0{\cdot}6 \times 10^{-11} \text{ V.-cm./(°C.)}^2,$$

$$b = +2{\cdot}2 \times 10^{-13} \text{ V.-cm.}^2\text{/(°C.)}^3,$$

$$e = +0{\cdot}0067\ \mu\text{V./°C.}$$

The same critical temperature, 79° C., is required for all tensions. Direct experimental tests with cooling-curves, a Paschen galvanometer being used to measure the thermo-e.m.f., verify this temperature exactly. All possible explanations of this transition are carefully discussed, the recrystallization theory appearing the most probable. Further work is suggested, and the significance of the present results for methods of increasing precision is pointed out.

§ 1. INTRODUCTION

In foregoing papers under the present general title analyses have been reported of measurements made on copper and nickel by Research Fellows of this Department. The analyses were successfully carried out in both cases on the assumption that copper and nickel were homogeneous metals at all temperatures under 200° C. In the case of aluminium, however, as reported in part III, the analysis was not successful, and the suggestion was there made that a transition temperature existed for aluminium below 180 C.

Here we shall report analyses of the results announced previously in this series, showing that this new suggestion permits of a very satisfactory explanation of all values of the e.m.f. obtained in the experimental work on aluminium.

§ 2. THEORETICAL BASIS

In the preliminary tests made by Li and Band to determine the most suitable sensitivity of the Paschen galvanometer the aluminium wire was heated to about the average temperature subsequently employed. It was found that several hours were necessary before a good steady state was attained. During this time different parts of the wire must have been at temperatures ranging between 25° and about 150° C., but the precise distribution of the temperature was not mapped.

The hypothesis which has been taken as the basis of the present analysis is simply that all those parts of the specimen which were, in this preliminary test, heated above a certain critical temperature T_c were changed in form and acquired in consequence a slightly different thermoelectric power from that of the original cold form of the metal. This change of form is supposed to be essentially a low-temperature recrystallization; it will therefore persist for a finite time after the temperature has decreased. Thus there are throughout the rest of the work two discontinuities in the wire at the points P and Q which were at the temperature T_c previously.

T_c

Whereas a homogeneous wire will exhibit an e.m.f. E given by the formula

E

$$E = a.A + b.B,$$

a, b

where A and B are the integrated values $\int G.dT$ and $\int G^2.dT$, the present hypothesis will modify this formula thus

A, B

$$E = a.A + b.B + e.DT,$$

where DT is the temperature-difference between the two points P and Q, and e is the difference between the thermoelectric powers of the two forms of metal.

DT

Neither the critical temperature T_c nor the positions of P and Q are known. But in analysis, we note that the maximum temperatures attained in the 1-kilogram tests were greater than those in either the preliminary tests or the zero-load tests. This means that the positions on the 1-kg. curves which are at T_c are the discontinuities for the 2-kg. tests. This gives a limiting factor in the choice of P and Q; in analysing the 2-kg. tests the choice of T_c fixes the positions of P and Q by reference to the 1-kg. tests.

The best method is by trial and error; possible values of T_c are chosen and from these the positions of P and Q and the values of DT are deduced.

If it is then possible to find values of a, b and e to give agreement between observed and calculated values of E, then we have a possible solution of the problem. Considering the variety of the temperature curves and values of e.m.f. observed, the discovery of even a single satisfactory solution would be a very convincing circumstance.

§ 3. RESULTS OF ANALYSIS

The integrals A and B were first worked out as described in the previous papers; they were tabulated with the observed values of E for all loads. Trial values of T_c were then selected, and the quantity DT was found for each temperature map occurring in the 2-kg. tests. The temperatures 130°, 120°, 110°, 100°, 90°, 80° and 75° were used; the positions of P and Q on the 1-kg. curve of highest temperature were then found, and from these the values of DT for each curve of the 2-kg. tests. These were tabulated along with the values of A, B and E.

Simple inspection was enough to eliminate most of these choices; only the last three gave approximate correlation, and of these, on more careful analysis, only 80° gave more than a few cases of good agreement.

For the next approximation, slight variations were made in the value of T_c, and hence in the positions of P and Q and in the values of DT. Owing to the steepness of the temperature-gradient at one of the points it is impossible to fix that point very precisely by the value of T_c; this permits an independent variation of DT by as much as a few degrees in most cases. It was thought that if within these latitudes a good correlation could be obtained, the choice of constants T_c, a, b, and e would be very effectively limited and therefore rather precisely determined.

Table 1 shows the results which were first worked out on the 2-kg. tests. The values of the constants are given in the third row of table 2, and the value of T_c was found to be 79° for best agreement.

Table 1. 2-kg. load

Curve	$A.10^{-3}$	$B.10^{-4}$	DT	E experimental (μV.)	E theoretical (μV.)
1	0·7	1·7	7	0·048	0·047
2	1·4	5·9	3	0·049	0·047
3	0·9	2·3	15	0·094	0·091
4	1·2	6·9	6	0·050	0·053
5	2·3	7·5	22	0·152	0·155
6	2·6	13·3	10	0·109	0·116
7	4·6	30·5	−6	0·103	0·104
8	4·3	21·8	31	0·265	0·263
9	4·3	28·3	27	0·263	0·263
10	5·9	48·0	0	0·181	0·203

In table 1 and those which follow, positive values of DT refer to cases where the point Q on the small temperature-gradient is at a higher temperature than the point P on the steep gradient.

With these values of T_c, a, b, and e of the 1-kg. and 3-kg. results, analysis was next attempted, and it was found that only reasonable changes in the latter three constants were required to give equally good agreement for the same value of T_c. The analysis of the results for zero tension have not quite as much significance, because the positions of P and Q are not fixed by T_c, no previous temperature-distribution having been mapped. P and Q were assumed to be the same as those used for the 1-kg. tests, and good enough agreement was then obtained. The values of the three constants are given in table 2, and the agreement is shown in the following tables.

Table 2

Load (kg.)	$a.10^{11}$	$b.10^{13}$	e (μV.)
0	+0·6	+2·2	+0·0067
1	0·5	2·0	0·0051
2	1·0	3·0	0·0050
3	1·2	3·4	0·0052

In every case, a change of one significant figure in the values given for the three constants destroys the good agreement, unless unreasonably large latitude or error is permitted in the positions of P and Q as determined by the critical temperature T_c.

Table 3. Zero load

Curve	$A.10^{-3}$	$B.10^{-4}$	DT	E experimental (μV.)	E theoretical (μV.)
1	0·3	0·65	4·5	0·032	0·033
2	3·3	20·50	−3·0	0·045	0·045
3	1·9	8·00	30·0	0·243	0·230
4	6·0	54·50	−13·0	0·066	0·068
5	5·1	41·00	23·0	0·253	0·254
6	3·1	18·60	16·0	0·156	0·155
7	5·0	38·00	8·0	0·165	0·167
8	3·8	23·00	13·0	0·159	0·161

Table 4. 1-kg. load

Curve	$A.10^{-3}$	$B.10^{-4}$	DT	E experimental (μV.)	E theoretical (μV.)
1	2·2	10·5	19·0	0·126	0·129
2	2·2	11·0	20·0	0·133	0·135
3	5·7	50·0	−6·0	0·097	0·099
4	1·7	7·8	17·0	0·119	0·112
5	3·6	23·0	4·0	0·085	0·085
6	3·5	24·0	−3·5	0·044	0·044
7	5·2	43·0	23·0	0·229	0·229

Table 5. 3-kg. load

Curve	$A.10^{-3}$	$B.10^{-4}$	DT	E experimental (μV.)	E theoretical (μV.)
1	1·3	5·0	4·0	0·054	0·054
2	2·2	12·0	3·5	0·085	0·084
3	3·2	21·0	−2·0	0·098	0·098
4	1·7	7·3	21·5	0·158	0·158
5	2·5	10·4	42·0	0·288	0·284
6	2·9	20·7	24·0	0·231	0·230
7	5·6	47·0	0·0	0·222	0·226

The load was not increased beyond 3 kg. because at 4 kg. the wire would approach closely to its elastic limit and the specimen became spoilt after a short time.

Positive values of e here obtained indicate that electrons flow in the recrystallized portion (between P and Q) from the hot end to the cold end—i.e. from Q to P, where DT is positive in the above tables. This indicates that the thermoelectric power of the recrystallized metal is greater than that of the normal metal by the value of e.

§4. EXPERIMENTAL TEST OF DISCONTINUITY AT 79° C.

The above indirect evidence of a discontinuity dependent upon temperature, which looks like a recrystallization of aluminium at 79°, permits of a simple experimental test. This was carried out as follows, after the foregoing analysis had been completed.

A new sample of the same aluminium wire was taken and mounted in much the same way as for the previous work, except that two thermocouples were attached to the wire at fixed points P and Q and a single travelling heater surrounded the wire between these points. The two copper-constantan couples were of the paper-insulated probe type previously used; they were adjacent to the specimen, and their junctions were bound to it with no. 40 s.w.g. copper wire tied around several times to improve thermal contact. Their free ends were connected with a galvanometer through a key so constructed that they were never connected electrically, except through the specimen.

The same Paschen galvanometer was used to detect the e.m.f. produced and, as before, full shielding was used to protect it against stray fields coming from the resistances controlling the heating-current. Sensitivity was set at about 100 mm./μV., at which value the zero is perfectly stable.

First the whole length of wire between P and Q was heated to about 70° C. for half an hour, and a test was made to see whether nearly symmetrical heating over P alone would produce an e.m.f. Temperature-differences of over 50° C. were tested between P and Q, and any discontinuity at these points of the kind postulated would have caused a measurable e.m.f. No indication of such an e.m.f. was found in this case except small deflections explicable as due to slight disturbances of the symmetry of temperature-distributions.

The process was repeated, P to Q being above 80° C. while P and Q were kept as nearly as possible at 80° C. This condition was maintained for about 15 minutes, and after a few minutes of cooling tests were again made at P and Q with the heater carrying only a small current. When P was at 60° and Q at 30° a definite voltage of the order of $0{\cdot}2\,\mu$V. was found. It vanished when the heater was placed symmetrically between P and Q, and which reversed sign when Q was heated and P allowed to cool.

Next, the heater was placed so that the thermocouple at Q just entered the end of the axis of the heater, and the couple P was shifted and remounted in a symmetrical position at the other end of the axis, say at P'. In this way the whole of $P'Q$ could be heated without moving the heater, while the original point P was not heated by more than a few degrees.

The current was turned on, and what must have been an almost symmetrical temperature-distribution caused a smoothly increasing e.m.f. as the temperatures at Q and P' rose. Both temperatures remained equal to within 0·5° C. The e.m.f. steadily increased until the temperatures reached 78°, when its value remained constant while the temperature rose to 84°. It then suddenly became irregular, increasing by a deflection of several millimetres and decreasing suddenly to the value which it had at 84°. These jerks continued until the temperature reached 90° when a steady increase again set in. At 95° the heating current was cut off and a curve of cooling was obtained; it was quite smooth except for a small lag at the commencement, and did not show any discontinuity at 80°; see *A* in figure 1.

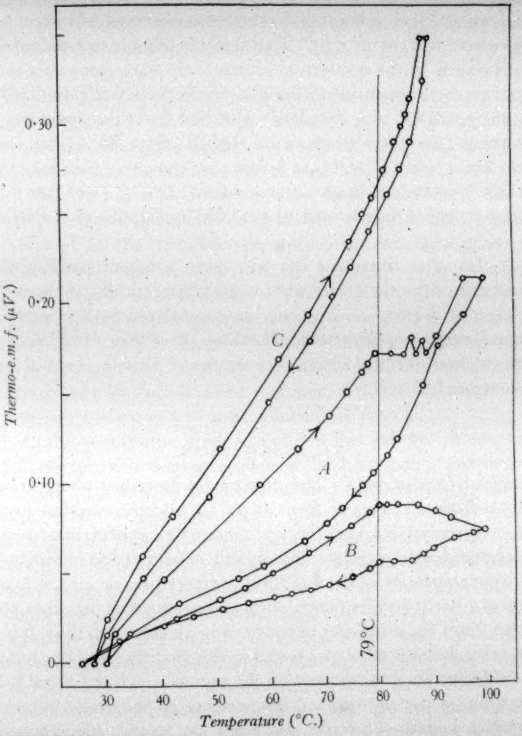


Figure 1.* Hysteresis cycles of thermo-e.m.f., showing the transition temperature at 79° C. (*a*) after 15 minutes local heating at 80° C.; (*b*) after 3 hours local heating at 80° C.; (*c*) after 6 hours entire heating at 80° C.

The part between *P′* and *Q* was now heated above 80° symmetrically for 3 hours, the two thermocouples being kept at about 80° throughout that time. After several hours of cooling the heater was shifted so that the point *Q* came to the middle of the axial bore in the heater, and the cycle shown at *B* in figure 1 was obtained. Here, as before, the heater carried a current sufficient to give a final temperature around 120°, and the rate of change of temperature with time was of convenient speed and quite steady. Temperature at the point *Q* and the e.m.f. were recorded every half-minute except at the last part of the cooling curve below 60°. It is to be noted that there was no increase in e.m.f. at temperatures above 79° where the steady value was reached; instead, a decrease was observed which continued when the heating current was cut off at 95°. That this was not due to a shift of the galvanometer zero is shown by the satisfactory return to the same zero after the apparatus had acquired room-temperature again. The whole cycle was completed in 1 hour.

---

* 12 hours later a cycle similar to *A* was obtained, showing partial recovery from recrystallization by the entire specimen.

Finally the specimen was completely removed from the apparatus, immersed in an oil bath and kept at a temperature slightly above 80° C. for 6 hours. Retesting at the same points *P′* and *Q* as before gave the curve *C* in figure 1; from the facts that there is no sudden break in the gradient at 79° C., and that the hysteresis loop has almost vanished, it is evident that the uniformity of the wire had been restored.

The fact that after treatment the wire gave a bigger e.m.f. than before is difficult to explain, since the heater was placed symmetrically between the thermocouples just as for the test *A*. The only directional physical property that could cause any e.m.f. under such symmetrical heating, other than a sudden discontinuity, would be an internal tortional strain. The results of Terado, Tsutsui and Tamano should be compared here(3).

§ 5. DISCUSSION

The change of form at 79° C. The direct tests described above are convincing evidence for a distinct change of form at 79° C., and the simplest explanation of this would be a recrystallization theory. Alternative explanations present themselves. A reaction between the aluminium and some metallic impurity in the wire may cause a heterogeneity at some critical temperature, or else an oxide coating may develop at a fairly definite temperature and produce its own thermo-e.m.f. In the latter case, since the oxide coat would become thicker with time, it is difficult to explain the quite definite value of *e* found in the analysis; also the temperature at which the oxide formed must depend on the rate at which the metal is brought to that temperature—a perfectly critical temperature is practically impossible.* The other alternative hypothesis would also seem too improbable, because chemical changes would be far too slow at such temperatures, and the reverse change would take place as easily on cooling; again, the actual temperature of transition would depend on the rate of change of temperatures and no true critical value would be possible.

The recrystallization theory seems therefore the most probable; but since commercial aluminium usually contains as much as 0·1 per cent of impurity,† it is just possible that recrystallization of a metallic impurity like aluminium carbide is responsible for the present results. It would obviously be desirable to repeat the direct test on 100-per cent pure metal, if such could be obtained in a practically useful form; but Peltier effects between junctions of the pure specimen with the less pure wire connexions would offer a considerable difficulty. Also, to avoid the criticism regarding oxidation, the specimen would have to be *in vacuo*, a provision which would make it almost impossible to avoid local heatings that would destroy the homogeneity of the specimen.

---

* It is also known that a very thin oxide film exists always at room-temperatures.

† Chiefly graphite and silica.

Methods of increasing precision. Assuming that our hypothesis is accepted as a working basis, probably the indirect analysis of the e.m.f. produced by mapped temperature-distributions is the most precise method of obtaining the change of thermoelectric power at recrystallization. It is almost impossible to get a perfectly symmetrical temperature-distribution so as to eliminate the e.m.f. due to the homogeneous effect and thus to isolate the exceedingly small e.m.f. due to contact between the two forms of the metal. Thus the direct test described in this paper, while precise enough for the critical temperature, is not capable of giving more than a rough estimate of e.

In order to get precise determinations of a and b, either all temperatures should be kept below 80°, which would make voltage-measurements more difficult, or else the whole specimen should be annealed at a temperature a little above 80° C. for several hours before each test. Preferably a and b should be found for both forms of the metal separately by both of these methods, and then e could be determined by using temperature maps in which both forms are present.

Throughout the quantitative work reported in this series of papers, reliance has been placed on the movable thermocouple and fixed junction in gentle contact with the specimen. The precision of the probe method has been discussed in part (I); but the present results for aluminium lend further support for that method.

Thus, it might be thought that a direct measurement of the temperature actually within the wire at several points would be a desideratum of precision; this has in fact been suggested to the writer by one competent critic. However, it is at once obvious that any attempt to use solder or welding to get real thermal contact between the junction and the specimen would heat the wire above the critical temperature here discovered, and so ruin the homogeneity of the specimen in an unmeasurable way. Further, any mechanical deformation of the wire required to insert a junction through its surface would again vitiate the claim that the wire is a homogeneous circuit, and would put our work back into the crude constriction class to which most previous work belongs(4).

It is therefore felt even more certainly that the only method of increasing precision is to increase the sensitivity of the voltage-measuring apparatus so that quite small temperatures and temperature-gradients can be used. By properly designed shields the space around the wire could be brought reasonably close to the temperature of the wire itself. While the probe obviously introduces a disturbance, no better method seems available; and such disturbance can be averaged out fairly completely in practice.

§ 6. CONCLUSION

In conclusion one may say that results for aluminium have been analysed and show excellently precise agreement with the homogeneous theory, if we also admit of a discontinuity of structure with a transition temperature at 79° C. The evidence that this discontinuity is due to the element aluminium is not perfectly conclusive until further tests have been made on 100 per cent pure metal. That the discontinuity is in the nature of a recrystallization seems the most probable explanation. It is obvious that ordinary methods of investigation such as recalescence during cooling would fail to detect such a slight change as appears to take place according to the present work. It is also doubtful if X-ray analysis could do so, because that method has given, until recently, somewhat inconsistent results when applied to the much greater recrystallization that occurs above 500° C.(5).

ADDITIONAL NOTE

by Dr W. BAND, received October 10, 1934

The three figures show typical results obtained on similar wire specimens. Symmetrical heating between two points A and B above 80° C. was maintained for several hours, no e.m.f. being observed. Immediately after cooling the heater was shifted over A, and a heating-cooling cycle of e.m.f. against T_A was obtained, figure 2. Immediately after this cycle had been completed, a second heating-cooling

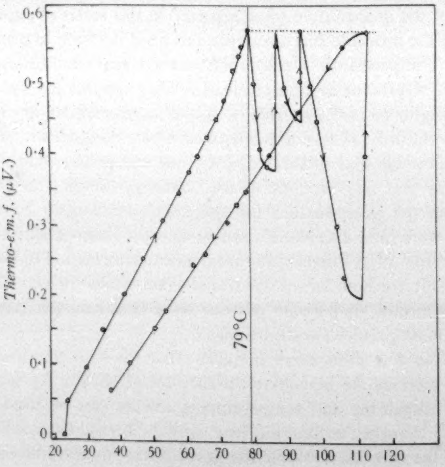


Figure 2.

process gave an entirely different curve above 79°, figure 3. Sample curves (e.m.f. against T_B) obtained when the heater was over B are shown in figure 4.

From these data we can see that the e.m.f. produced in the newly recrystallized parts of the wire depends for its direction upon the direction of the temperature-gradient in those parts, viz.: the difference between figure 2 and figure 4. But it is difficult to explain how a curve like that of figure 3 could follow the cycle shown in figure 2.

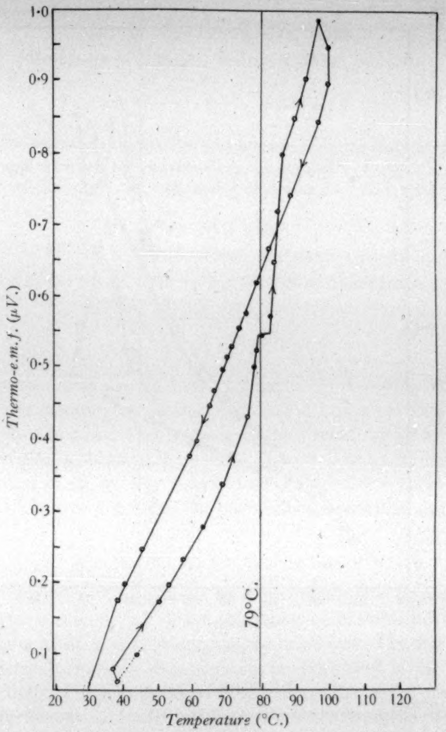


Figure 3.

The magnitudes of the voltages obtained in these tests seem to depend upon the mechanical treatment of the wire during mounting in the apparatus, and thus presumably upon small internal strains. Such strains would only even out after a considerable time. This would explain why, during the original measurements previously reported, it was found necessary to maintain a steady temperature-distribution for 4 hours or more before a steady e.m.f. could sometimes be obtained.

That the so-called transition at 79° C. may be connected with slight re-adjustments of internal strain rather than with recrystallization is an interesting possibility. It is felt, however, that further speculation had better await the results of more refined methods of investigation suggested by the experiences of the present exploratory work.

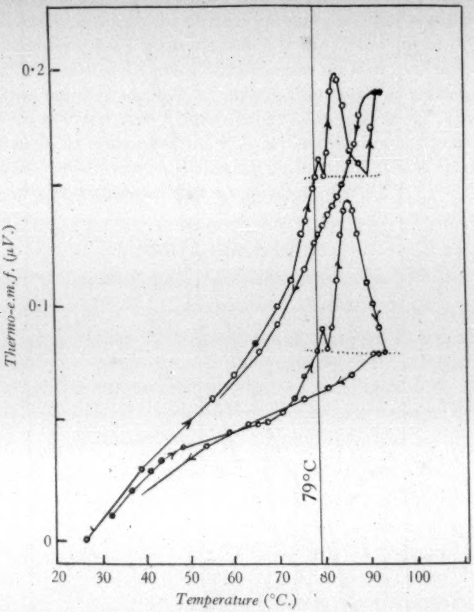


Figure 4.

REFERENCES

(1) Li and Band. The longitudinal thermoelectric effect: (3) aluminium, p. 859 of this volume.

(2) Li and Band. The longitudinal thermoelectric effect: (1) copper. *Proc. phys. Soc.* **46**, 515 (1934).

(3) Terado, Tsutsui and Tamano. *Sci. Pap. Inst. phys. chem. Res.*, Tokyo, **7**, Nos. 122 and 195.

In this connexion one would like to be more certain that a symmetrical temperature-distribution can be moved along a wire and still remain symmetrical; temperature waves are almost sure to commence with the motion of the heater, and these would at once cause the e.m.f. fluctuations observed by these workers. There seems no particular need to postulate fluctuations of internal crystal-arrangement.

(4) W. Hume-Rothery, in his *The Metallic State*, pp. 107–114 (Oxford 1931), gives an excellent critical summary of work done previously to that date.

(5) Schmid and Wassermann. *Metallwirtschaft*, **10**, 409–10 (1931).

539.17

THE DISINTEGRATION OF BORON BY NEUTRONS

By H. J. TAYLOR, M.Sc.
Professor of Physics, Wilson College, Bombay

Received May 11, 1935. *Read June* 7, 1935

ABSTRACT. The tracks of the particles which result from the disintegration of boron by neutrons have been recorded by impregnating the emulsion of a photographic plate with borax. It is shown that there are two modes of disintegration, corresponding to the reactions:

$$B^{10} + n^1 \rightarrow Li^7 + He^4,$$
$$B^{10} + n^1 \rightarrow He^4 + He^4 + H^3.$$

Photomicrographs of the tracks, showing both modes of disintegration, are reproduced.

§ 1. INTRODUCTION

A BRIEF report has already been published(1), recording the discovery of tracks corresponding to the disintegration of boron and lithium when bombarded by neutrons. The purpose of the present paper is to give a fuller statement of the results obtained in the case of boron. The tracks of the fast particles which are liberated in the nuclear reactions are recorded directly in a photographic emulsion. The technique employed has been fully described in a previous paper(2).

§ 2. METHOD OF EXPERIMENT

Briefly stated, the method consists in impregnating the emulsion of a photographic plate with a suitable salt of the element to be examined. The plate is then exposed to the neutrons, and developed in the usual way. The tracks of the heavy particles which result from the disintegration are registered in the emulsion; each track is visible, under a high magnification, as a row of developed grains in a straight line. For satisfactory work special emulsions free from background and of very fine grain are required. Ilford R plates have been specially prepared to meet these requirements, and have been used for the present experiments.

The source of neutrons for most of the experiments consisted of a quantity of radon sealed in a tube with powdered beryllium. The initial amount of radon varied for different experiments, the maximum being about 200 millicuries. Some experiments were also done with a polonium-beryllium source, but this was found less satisfactory on account of the comparative weakness of the polonium sources available. Exposure to fast neutrons was effected by placing the plate as near as possible to the source, but with some 6 cm. of lead interposed, in order to reduce the direct effect of the γ rays on the plate. To expose the plate to slow neutrons, both source and plate were surrounded by a large quantity of paraffin. Under these conditions, however, fast neutrons are still present in the radiation reaching the plate. Conversely, some slow neutrons are present even in the absence of paraffin, on account of scattering which takes place in the table and other surrounding objects.

§ 3. THE DISINTEGRATION OF BORON

Boron is introduced into the emulsion by soaking the plate in a strong solution of borax and allowing it to dry in the dark. It is only possible to make a rough estimate of the amount of boron thus introduced, but it is of the order of 0·05 mg. per cm^2 of plate-surface. This is only 1 per cent of the amount required, for slow neutrons, to reduce the intensity of the radiation to one half of its initial value(3). The treatment of the plate with borax was found not to affect the photographic properties of the emulsion seriously. The borax was generally washed out before development, but this seems unimportant.

Disintegration by slow neutrons. With slow neutrons the probability of disintegration is relatively high, and exposures of the order 1000 millicurie-hours were found to be ample. The best plates show as many as 50,000 tracks per cm^2, sensibly straight, and with the number of grains per track varying up to ten. The mean length of the tracks is $7{\cdot}6\,\mu$. Photomicrographs of typical tracks are shown in figures 1 and 2. Most of the tracks are very clear and well defined, but it is difficult to determine the equivalent range in air exactly, for the following reasons. (i) The measured length (i.e., the distance between the extreme grains of a track) shows, as one would expect, large fluctuations owing to the random distribution of grains in the emulsion. (ii) The true range will in general exceed the measured length by an unknown amount, since the end grains need not be at the extreme points of the range. (iii) The conversion factor to reduce the range in gelatine to the range in air is subject to some uncertainty. The uncertainties can, however, be partly allowed for on the basis of previous experience(2), and the best value for the equivalent range in air which can at present be deduced from the measurements is

$$1{\cdot}1 \pm 0{\cdot}1 \text{ cm.}$$

These tracks must be interpreted as being due to the disintegration of the boron nucleus into two particles. As the momentum of the neutron is small these particles move in sensibly opposite directions, giving rise to an apparently single track.

With our present knowledge of nuclear masses, the only reaction which appears to fit these observations is

$$\text{B}^{10} + n^1 \rightarrow \text{Li}^7 + \text{He}^4 \qquad \ldots\ldots(1).$$

The release of energy, corresponding to the observed range, is 2×10^6 electron-volts.

Until recently, the accepted masses of the neutral atoms of Li^7 and He^4 and of the neutron have been as follows: $\text{Li}^7 = 7{\cdot}0146$, $\text{He}^4 = 4{\cdot}0022$, $n^1 = 1{\cdot}0080$. With these values reaction (1) requires the value $\text{B}^{10} = 10{\cdot}011$, disagreeing markedly with Aston's value $\text{B}^{10} = 10{\cdot}0135$.

Recently, however, the available evidence has been reviewed(4, 5, 6), and it

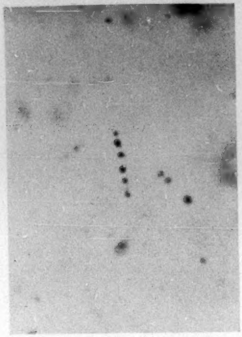


Figure 1.

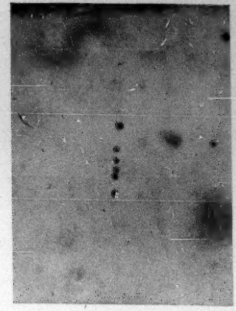


Figure 2.

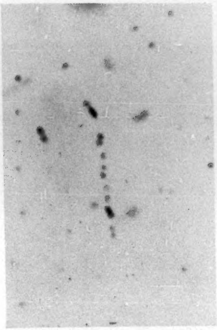


Figure 3.

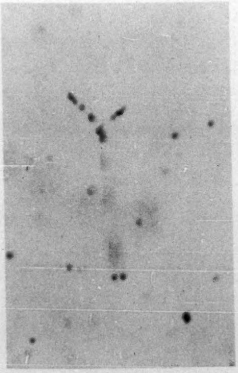


Figure 4.

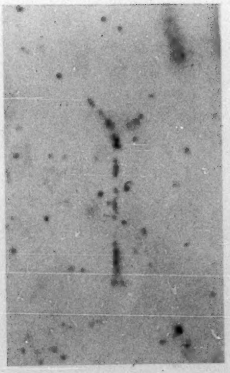


Figure 5.

(1) and (2) Typical examples of tracks resulting from reaction (1), and due to a slow neutron.
(3) Tracks corresponding to reaction (1), but due to a fast neutron.
(4) Three tracks corresponding to reaction (2). The long H^3 track rises out of the plane of focus.
(5) The same photographed in a higher plane, showing the H^3 track to better advantage.
Magnification of all the photographs: × 1930.

appears that the accepted masses may require some modification. At the moment the most probable values are$^{(4)}$

$$n^1 = 1{\cdot}0083$$
$$He^4 = 4{\cdot}0034$$
$$Li^7 = 7{\cdot}0170$$
$$B^{10} = 10{\cdot}0143.$$

These masses fit reaction (1) within the limits of error.

Disintegration by fast neutrons. Boron under fast neutron bombardment gives many disintegrations of type (1), some of which are due to the slow neutrons always present in the radiation. In some cases the tracks of Li^7 and He^4 can be separately seen, as the angle between them is less than 180°. This is an indication that a fast neutron, with considerable momentum, has been responsible for the disintegration. Figure 3 shows an example of this.

Several examples of what appears to be a different type of disintegration have been found. The best example is reproduced in figures 4 and 5. Three tracks diverge from a single point, two shorter ones denoted by a and b, and a longer one denoted by c. Track c does not lie in the plane of the photograph, but rises steeply from the point of disintegration. The horizontal component of c may be denoted by c'.

The ranges and angles are as follows:

$a = 0{\cdot}7$ cm. in air, angle $a, b = 102°$
$b = 0{\cdot}5$ cm. in air, angle $b, c' = 120°$
$c = 3{\cdot}0$ cm. in air, angle $c', a = 138°$
angle $c, c' = 35°$.

The angle c–c' in the vertical plane is subject to an uncertainty of about 5°.

One might advance the hypothesis that two of the tracks, a and c, represent a disintegration of type (1), whilst the third track is independent and situated at this point purely by chance. On this hypothesis the resultant momentum of the two particles can be worked out, and this must be the same as the momentum of the incident neutron. The energy of the incident neutron calculated in this way is found to be 13×10^6 eV., of which half must have been absorbed in the reaction. It is evident that the hypothesis breaks down, even when allowance is made for very considerable errors in the measurements. There can thus be little doubt that we are really dealing here with a disintegration into three particles. It appears that the only reaction which agrees with the observations is

$$B^{10} + n^1 \rightarrow He^4 + He^4 + H^3 \qquad \text{......(2)}.$$

With the masses recently proposed$^{(4)}$ ($H^3 = 3{\cdot}0161$), it will be seen that the reaction requires an energy of 3×10^5 eV. to be supplied, which explains why it does not occur with slow neutrons. In this example the combined energy of the three particles is $3{\cdot}3 \times 10^6$ eV., so that the energy of the incident neutron must have been $3{\cdot}6 \times 10^6$ eV.

The resultant momentum of the three particles can be calculated, but the uncertainty in the angle c–c' affects the value obtained. This resultant momentum should be the same as that of the incident neutron, and within the limits of accuracy of the measurements no discrepancy is found.

Several other examples of what are apparently similar cases of triple disintegration have been found, but no other which is so well placed for measurement. This type of disintegration is clearly much rarer than type (1), the relative frequencies of occurrence, in the plates exposed to fast neutrons, being of the order 1 : 1000.

§ 4. ACKNOWLEDGMENTS

My thanks are due to Mr Goldhaber, who suggested some months ago that lithium and boron might be amenable to investigation by this method, and who has also helped in making the exposures. I also desire to thank Lord Rutherford and Dr Chadwick of the Cavendish Laboratory, where the experiments have been done, for their interest and encouragement. I am greatly indebted to Mr Bloch and his staff, of the Ilford Company, for producing the special photographic material. Finally, I must express my thanks to the Senatus of Wilson College, Bombay, for extended leave of absence.

REFERENCES

(1) TAYLOR and GOLDHABER. *Nature*, Lond., **135**, 341 (March 2, 1935).
(2) TAYLOR. *Proc. roy. Soc.* **150**, 382 (1935).
(3) AMALDI, D'AGOSTINO, FERMI, PONTECORVO, RASETTI, and SEGRÉ. *Proc. roy. Soc.* **149**, 522 (April 10, 1935).
(4) OLIPHANT, KEMPTON, and RUTHERFORD. *Proc. roy. Soc.* **150**, 241 (May 1, 1935).
(5) BETHE. *Phys. Rev.* **47**, 633 (April 15, 1935).
(6) ASTON. *Nature*, Lond., **149**, 522 (April 10, 1935).

535·375·54:538.221

A NOTE ON THE RAMAN SPECTRUM OF A FERROMAGNETIC OXIDE

By L. F. BATES, D.Sc., Ph.D., Reader in Physics,
University College, London

AND

H. E. HOGWOOD, B.Sc.

Received April 4, 1935.

ABSTRACT: An attempt has been made to determine whether the ferromagnetic oxide Cr_5O_9 exhibits a Raman spectrum. No trace of such a spectrum was found.

In recent years our knowledge of the ferromagnetism of metals has been considerably increased, but our knowledge of the nature of ferromagnetism in the metallic oxides is still very meagre. Now, it occurred to one of us (L. F. B.) that valuable information might be obtained if it were possible to examine the Raman spectra of such oxides, for changes in the Raman spectrum with rise in temperature might show whether a particular grouping of atoms was responsible for the observed ferromagnetic properties, and might also indicate the nature of the coupling between the atoms. A perusal of the literature showed that the oxide of chromium, Cr_5O_9, would probably be the best substance with which to make an attempt. The magnetic properties of this oxide have been studied by Dr H. Sachse(1), who found that it possesses a ferromagnetic Curie point at about 108° C. We are deeply indebted to Dr Sachse and to Fräulein Lorentz for the gift of samples of Cr_5O_9, which we were able to compare with those prepared for our work.

The oxide, Cr_5O_9, was prepared by passing a stream of chromyl chloride vapour over hot glass or silica maintained at a temperature of about 370° C. Suitable heavy deposits were usually obtained in a few hours, although thin, translucent deposits were formed on the inner walls of tubes in a few minutes. The preparation of crystals of Cr_5O_9 as large as those obtained by Dr Sachse requires much time and patience, especially as glass becomes brittle after being in contact with hot chromyl chloride vapour for some time. The oxide shows a pronounced violet surface colour and appears green by transmitted light. It is difficult to determine the thickness of a layer unless it is deposited on a previously weighed piece of glass, for it is extremely difficult to remove.

In a preliminary survey the light from a powerful quartz mercury vapour lamp was concentrated upon and passed through a thin oxide deposit formed on the inner surface of a bulb of transparent silica, about 2 cm. in diameter, so as to enter a quartz Raman spectrograph of great light-collecting power, manufactured by Messrs Hilger. Direct light was carefully excluded, and with an exposure of 150 hours it was found that little beyond the 3132 lines was transmitted, and no new lines appeared.

A more thorough survey, particularly in the ultra-violet, was made by concentrating the light upon a thick coating of the oxide inside a tube which had been cut obliquely at one end, the light falling upon the projecting portion. The light was then scattered from the deposit many times before it reached the collimator slit. The result of a 100-hour exposure to the scattered light was compared with that of a 10-sec. exposure to the direct beam. The chief difference between the two spectrograms was the appearance of some fluorescence on the long-wave sides of the more intense lines, but no evidence of new or Raman lines was found.

This method of attack was less satisfactory for the visible region of the spectrum, where the following procedure was adopted. Heavy deposits of Cr_5O_9 were made on two strips of fused silica 12 cm. long and 2·5 cm. wide. These were mounted parallel, with the deposits facing each other, but so displaced laterally that the incident light could be concentrated upon the one plate. After being scattered several times, the light was finally scattered from the other plate into the spectrograph. The result of 300 hours' exposure with this arrangement, when compared with that obtained with one hour's exposure with the deposits covered with white blotting paper, again showed no Raman lines to be present.

It is therefore concluded that no Raman lines can be obtained in experiments on light scattered from the upper layers of this ferromagnetic oxide. It must be recorded, however, that the experiments here described were made under very adverse conditions of scattering, and the main purpose of this note is to report the attempts made, in the hope that other workers may possibly discover a ferromagnetic compound more suitable for Raman investigations.

In conclusion, we desire to thank Professor E. N. da C. Andrade for the facilities placed at our disposal, especially for the loan of the Raman spectrograph, and Mr H. Terrey for help in preparing the oxide deposits.

REFERENCE

(1) H. Sachse, *Zeit. für Phys.* **70**, 546 (1931).

548.73

THE ABSORPTION FACTOR FOR THE POWDER AND ROTATING-CRYSTAL METHODS OF X-RAY CRYSTAL ANALYSIS

By A. J. BRADLEY, D.Sc., Royal Society Warren Research Fellow

Communicated by Prof. W. L. Bragg, F.R.S., June 17, 1935.

ABSTRACT. A method has been devised for calculating the absorption factor for the powder method. It may be used for all values of μr, where μ is the linear absorption coefficient and r the radius of the specimen. It is, however, especially suitable when μr exceeds 2. For smaller values of μr, for which the calculations are tedious and become less accurate, it is better to use Claassen's graphical method. By such a combination of the two methods a set of data was calculated for which the absorption factor may easily be obtained for any value of μr and for any angle of reflection, with an error not exceeding 1 per cent. The data may also be applied for the calculation of the absorption factor in the lines belonging to the layer of zero order in rotation photographs.

---

§ 1. INTRODUCTION

In the Debye-Scherrer method of crystal analysis a beam of X rays bathing a cylindrical specimen of powdered crystalline material gives rise to a diffraction pattern. The relative intensities of the lines in this pattern are very much influenced by the absorption of the X rays in the specimen, but hitherto no entirely satisfactory method has been devised for calculating the numerical value of the absorption factor. Solutions for particular cases have been given by different authors$^{(1, 2, 3)}$, while Claassen$^{(4)}$ has devised a graphical method which is of general application. Rusterholz$^{(5)}$ has shown that for very highly absorbing specimens it is possible to calculate the absorption factor, which is in this instance a fairly simple function of the glancing angle θ. θ

It is, however, possible to calculate the approximate absorption factor by a new method which gives results accurate to at least 1 per cent, except for very small absorptions where the Claassen graphical method is entirely satisfactory. The chief importance of the new method for calculating the absorption factor is that it gives accurate results in the region where Claassen's method is less reliable. Results are here tabulated for all values of μr, where μ is the linear absorption coefficient of the μ specimen and r is its radius. r Only a limited number of angles have been considered, but the absorption factor for intermediate angles is quickly found by graphical interpolation.

For either a single crystal or a mass of crystalline powder the intensity of reflection, in the absence of absorption, is proportional to the total volume V of the V

h μ a $d\sigma$

specimen which is immersed in the incident X-ray beam. Thus for a cylindrical specimen the intensity is proportional to $\pi r^2 h$, where h is the height of the specimen in the beam. This statement does not hold good where there is finite absorption of X rays in the specimen, the contribution of different portions of the specimen being no longer equivalent. In many cases the absorption effects are so great that only a small portion of the periphery of the specimen makes any appreciable contribution towards the intensity of reflection. The contribution of a tiny fragment of crystal of volume dV is reduced by the fraction $e^{-\mu a}$, where μ is the linear absorption coefficient of the whole specimen and a is the total length of path of the ray through the specimen before and after reflection from the fragment. In order to evaluate the total strength of the reflection, the value of $e^{-\mu a}\,dV$ must be integrated throughout the volume of the specimen. The resulting expression then replaces the factor V in the intensity formula. For a cylindrical specimen $\pi r^2 h$ is replaced by the expression $h\iint e^{-\mu a}\,d\sigma$, where $d\sigma$ is a small element of the cross-section of the specimen, and the integration is carried out over the whole cross-section. The intensity of the reflection is, therefore, cut down by absorption in the ratio

A

$$A=\frac{1}{\pi r^2}\iint e^{-\mu a}\,d\sigma.$$

A is termed the absorption factor. It may be expressed in the alternative form

$$A=\iint e^{-\mu r x}\,ds \qquad \text{......(1)},$$

x

where $x=a/r$ and $ds=d\sigma/\pi r^2$.

s

To evaluate equation (1) it is necessary to obtain the value of ds/dx for all possible values of x. The first step (described in § 2) is to calculate the value of s, the proportion of the cross-section of the specimen for which the absorbing path xr is less than a given value—e.g. the area $XSX'Y$ in figure 1. This gives s as a function of x, which may be differentiated to give ds/dx. The problem is made more complicated because the same formulae are not applicable to all portions of the cross-section of the specimen. In figures 3 to 8 different portions are distinguished according to the formulae which are applicable. The arguments used in § 2 are based primarily on figure 1, which is typical of the front areas of figures 3 to 8. Figure 2 is typical of a back area. No direct method could be found for calculating the corresponding formulae for the centre areas. Suitable corrections can, however, be made for the latter. Except for small absorption, the formulae deduced for the front areas are the most important, and for large values of μr the rest may be disregarded.

The calculations in § 2 give s as an ascending-power series in x, which is convergent for $x \not> 2$. This is quite satisfactory since, as will be seen later, x cannot exceed 2 in the front areas of figures 3 to 8. Claassen has suggested that such a series might be used to calculate the absorption factor for large values of μr in the following way.

Putting

$$s=\alpha x+\beta x^2+\gamma x^3+\dots \qquad \text{......(2)},$$

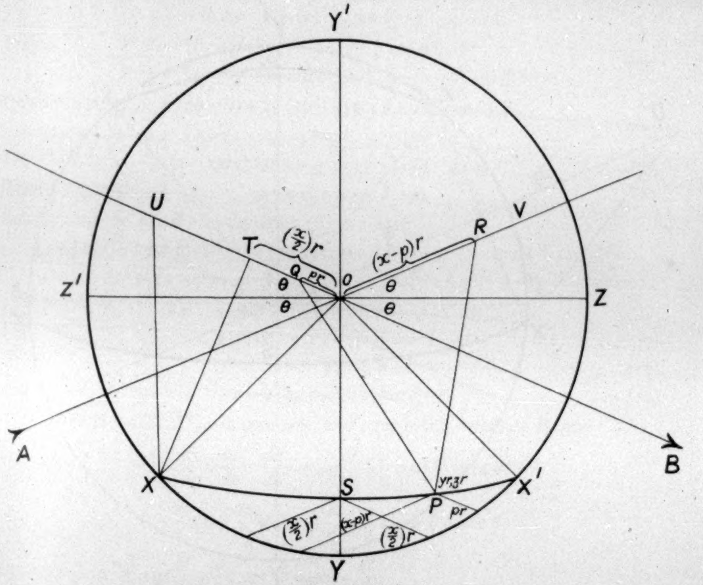
we find that

$$\frac{ds}{dx}=\alpha+2\beta x+3\gamma x^2+\dots \qquad \text{......(3)}.$$

Then according to Claassen equation (1) for large values of μr gives approximately

$$A=\int_{x=0}^{x=\infty}\frac{ds}{dx}\,e^{-\mu r x}\,dx=\frac{\alpha}{\mu r}+\frac{2!\,\beta}{\mu^2 r^2}+\frac{3!\,\gamma}{\mu^3 r^3}+\dots \qquad \text{......(4).}$$

It will be found, however, that if one substitutes for α, β, γ, etc. the coefficients obtained in § 2, the above series becomes divergent for all values of μr and is therefore invalid. This difficulty is avoided by replacing the upper limit $x=\infty$ by the exact maximum value of x for the front portion of the specimen. Then, unfortunately, equation (4) becomes considerably more complicated, and since the maximum value of x is to some extent a function of θ it is necessary to consider separately the cases of reflection at different angles.



General case

Figure 1.

The value of A found from a series such as that in equation (2) represents the contribution of either the front or the back portion of the specimen. For the centre portions an estimate must suffice, since it is not possible to obtain a series for them. The complete absorption factor is the sum of the values for the front, back and centre portions. However, the contribution from the front portion usually far outweighs the others, which are really in the nature of correction terms.

§ 2. THE AREA OF THE SPECIMEN FOR WHICH THE ABSORBING PATH IS LESS THAN A GIVEN VALUE

In figure 1 AO and OB indicate respectively the directions of the incident and reflected beams. The incident beam is supposed to be parallel and completely to Y, Z bathe a cylindrical specimen of radius r. θ is the glancing angle. YOY' and ZOZ' X are respectively normal and parallel to the reflecting crystal planes. XSX' is the locus of all reflecting points for which the sum of the incident and emergent paths

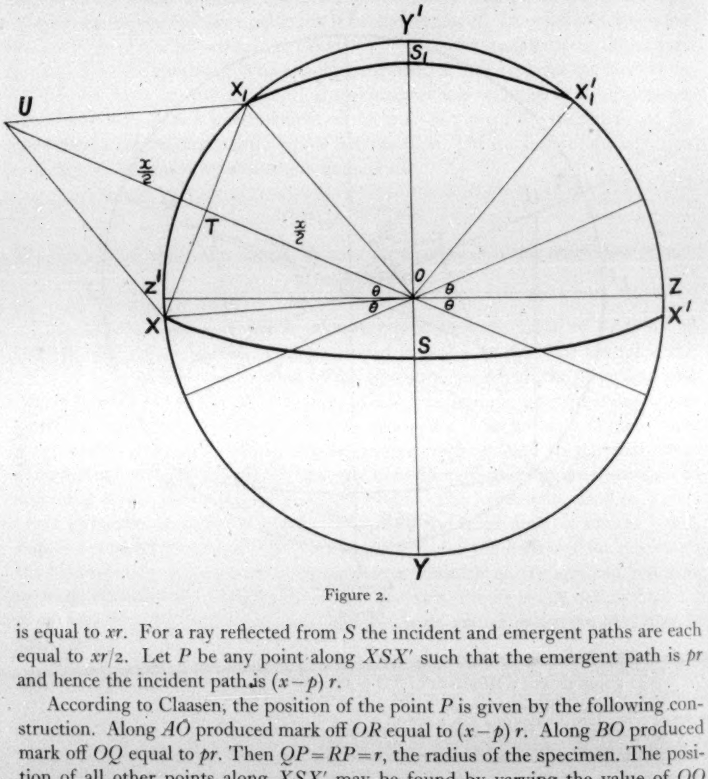


Figure 2.

is equal to xr. For a ray reflected from S the incident and emergent paths are each p equal to $xr/2$. Let P be any point along XSX' such that the emergent path is pr and hence the incident path is $(x-p)\,r$.

According to Claasen, the position of the point P is given by the following construction. Along AO produced mark off OR equal to $(x-p)\,r$. Along BO produced mark off OQ equal to pr. Then $QP=RP=r$, the radius of the specimen. The position of all other points along XSX' may be found by varying the value of OQ between zero and xr, while that of OR is varied between xr and zero.

It is required to find the area of $XSX'Y$, which is that portion of the specimen in which all particles are situated for which the sum of the incident and reflected paths is less than x. It is first necessary to determine each of the areas $OXYX'$ and $OXSX'$ separately.

The area
$$OXSX' = r^2 \int_{z=0}^{y^2+z^2=r^2} y\,dz - r^2 \int_{z=0}^{y^2+z^2=r^2} z\,dy \qquad \text{......(5)}.$$

It is necessary to evaluate y and z in terms of a common parameter, for which purpose the following relation is used:

$$PQ = PR = r.$$

The y, z coordinates of Q are $\quad -p\sin\theta,\ -p\cos\theta.$

R are $\quad -(x-p)\sin\theta,\ (x-p)\cos\theta.$

Hence
$$\left.\begin{aligned} PQ^2 &= r^2(y+p\sin\theta)^2 + r^2(z+p\cos\theta)^2 = r^2 \\ PR^2 &= r^2\{y+(x-p)\sin\theta\}^2 + r^2\{z-(x-p)\cos\theta\}^2 = r^2 \end{aligned}\right\} \qquad \text{......(6)}.$$

On subtraction,
$$x(2p-x) + 2y(2p-x)\sin\theta + 2zx\cos\theta = 0,$$
$$z = (1-2p/x)\tan\theta\,(\tfrac{1}{2}x\operatorname{cosec}\theta + y).$$

Let
$$(1-2p/x)\tan\theta = q \quad\text{or}\quad p = \tfrac{1}{2}x(1-q\cot\theta) \qquad \text{......(7)}.$$

Then
$$z = q(\tfrac{1}{2}x\operatorname{cosec}\theta + y) \qquad \text{......(8)}.$$

Substituting for z and p in equation (6), we have

$$\{y+\tfrac{1}{2}x(1-q\cot\theta)\sin\theta\}^2 + \{q(\tfrac{1}{2}x\operatorname{cosec}\theta+y) + \tfrac{1}{2}x(1-q\cot\theta)\cos\theta\}^2 = 1,$$
$$y^2(1+q^2) + xy(1+q^2)\sin\theta + (\tfrac{1}{2}x)^2(1+q^2) - 1 = 0,$$
$$y = -\tfrac{1}{2}x\sin\theta \pm \sqrt{\{1/(1+q^2) - (\tfrac{1}{2}x)^2\cos^2\theta\}}.$$

Let
$$\pm\sqrt{\{1/(1+q^2) - (\tfrac{1}{2}x)^2\cos^2\theta\}} = t \qquad \text{......(9)}.$$

Then
$$y = -\tfrac{1}{2}x\sin\theta + t \qquad \text{......(10)},$$

and
$$z = q(\tfrac{1}{2}x\cos^2\theta\operatorname{cosec}\theta + t) \qquad \text{......(11)}.$$

For the locus XSX', x is constant, and p (q) is the variable. Hence

$$\frac{dy}{dq} = \frac{dt}{dq};\quad \frac{dz}{dq} = t + q\frac{dt}{dq} + \frac{x}{2}\cos^2\theta\operatorname{cosec}\theta,$$

$$y\frac{dz}{dq} = -\frac{x}{2}\sin\theta\frac{dz}{dq} + t^2 + qt\frac{dt}{dq} + \frac{x}{2}t\cos^2\theta\operatorname{cosec}\theta,$$

$$z\frac{dy}{dq} = qt\frac{dt}{dq} + \frac{x}{2}q\cos^2\theta\operatorname{cosec}\theta\frac{dt}{dq},$$

$$\int y\,dz = -\tfrac{1}{2}x\sin\theta\int dz + \int t^2dq + \int qt\,dt + \tfrac{1}{2}x\cos^2\theta\operatorname{cosec}\theta\int t\,dq,$$

$$\int z\,dy = \int qt\,dt + \tfrac{1}{2}x\cos^2\theta\operatorname{cosec}\theta\int q\,dt.$$

The required area may now be found from equation (5), inserting the appropriate limits. For the case shown in figure 1, we have

$$\text{at } S,\ p = \tfrac{1}{2}x,\ q = 0;\ \text{ at } X',\ p = 0,\ q = \tan\theta.$$

Hence
$$\text{Area } OXSX' = r^2\int_{q=0}^{q=\tan\theta}(y\,dz - z\,dy)$$
$$= r^2\left[-\tfrac{1}{2}x\sin\theta\int_{q=0}^{q=\tan\theta}dz + \int_{q=0}^{q=\tan\theta}t^2dq + \tfrac{1}{2}x\cos^2\theta\operatorname{cosec}\theta\int_{q=0}^{q=\tan\theta}(t\,dq - q\,dt)\right] \qquad \text{......(12)}.$$

The above integral can be expanded from equations (9), (10), (11). The ambiguity in equation (9) leads to two solutions. Since figure 1 shows a positive value for y, the positive sign must be taken for t. This rule applies throughout the front areas of figures 3 to 9. For the back areas of figures 2, 3 and 4, y is always negative. These areas correspond to negative t in equation (9). The two cases will be considered separately, and finally a reference will be made to the centre areas of figures 3 to 8 which do not fall into either class.

Case 1. *Front Areas only, as in figure* 1. From equations (9), (10), (11), (12), taking the positive sign in equation (9),

$$\text{Area } OXSX' = r^2\Big[-(\tfrac{1}{2}x)^2 \sin 2\theta - \tfrac{1}{2}x\sqrt{\{1-(\tfrac{1}{2}x)^2\}} + \theta$$
$$+ x\cos^2\theta \operatorname{cosec}\theta \int_{q=0}^{q=\tan\theta} \sqrt{\{1/(1+q^2) - (\tfrac{1}{2}x)^2\cos^2\theta\}}\, dq\Big] \quad \ldots\ldots(13a).$$

Area $OXYX'$ is found from figure 1 as follows:

Let $$OU = xr;\ OT = \tfrac{1}{2}xr.$$

Then X, the end of the locus XSX', is given by

$$UX = OX = r.$$

Since T is the midpoint of OU, XTO is a right-angled triangle. Hence

$$\cos T\hat{O}X = \tfrac{1}{2}x,$$

$$X\hat{O}S = \pi/2 - X\hat{O}Z' = \pi/2 - X\hat{O}T + Z'\hat{O}T = \sin^{-1}(\tfrac{1}{2}x) + \theta.$$

Hence area $$OXYX' = r^2 X\hat{O}S = r^2\{\sin^{-1}(\tfrac{1}{2}x) + \theta\} \quad \ldots\ldots(14a).$$

From equations (13a) and (14a), in figure 1

$$\text{Area } XSX'Y = \text{area } OXYX' - \text{area } OXSX'$$
$$= r^2\Big[\sin^{-1}(\tfrac{1}{2}x) + (\tfrac{1}{2}x)^2\sin 2\theta + \tfrac{1}{2}x\sqrt{\{1-(\tfrac{1}{2}x)^2\}} - x\cos^2\theta\operatorname{cosec}\theta$$
$$\times \int_{q=0}^{q=\tan\theta} \sqrt{\{1/(1+q^2) - (\tfrac{1}{2}x)^2\cos^2\theta\}}\, dq\Big] \quad \ldots\ldots(15a).$$

Case 2. *Back Areas only, as in figure* 2. From equations (9), (10), (11), (12), taking the negative sign in equation (9),

$$\text{Area } OX_1S_1X_1' = r^2\Big[-(\tfrac{1}{2}x)^2\sin 2\theta + \tfrac{1}{2}x\sqrt{\{1-(\tfrac{1}{2}x)^2\}} + \theta$$
$$- x\cos^2\theta\operatorname{cosec}\theta \int_{q=0}^{q=\tan\theta} \sqrt{\{1/(1+q^2) - (\tfrac{1}{2}x)^2\cos^2\theta\}}\, dq\Big] \quad \ldots\ldots(13b).$$

This formula gives a negative value for the area, because it lies completely on the negative side of the z axis.

Area $OX_1Y'X_1'$ from figure 2 by the same construction as that used in figure 1 gives

$$X_1\hat{O}S_1 = \pi/2 - X_1\hat{O}Z' = \pi/2 - X_1\hat{O}T - Z'\hat{O}T = \sin^{-1}(\tfrac{1}{2}x) - \theta.$$

Hence area $$OX_1Y'X_1' = r^2 X_1\hat{O}S_1 = r^2\{\sin^{-1}(\tfrac{1}{2}x) - \theta\} \quad \ldots\ldots(14b).$$

From equations (13b) and (14b) we may deduce the area $X_1S_1X_1'Y'$ in figure 2.

Since the area given by equation (13*b*) is a negative quantity, it is necessary to change the signs before subtraction.

$$\text{Area } X_1S_1X_1'Y' = \text{area } OX_1Y'X_1' - \text{area } OX_1S_1X_1'$$

$$= r^2\left[\sin^{-1}(\tfrac{1}{2}x) - (\tfrac{1}{2}x)^2\sin 2\theta + \tfrac{1}{2}x\sqrt{\{1-(\tfrac{1}{2}x)^2\}} - x\cos^2\theta\operatorname{cosec}\theta \times \int_{q=0}^{q=\tan\theta}\sqrt{\{1/(1+q^2)-(\tfrac{1}{2}x)^2\cos^2\theta\}}\,dq\right] \quad \text{......(15}b\text{)}.$$

It will be seen that equations (15*a*) and (15*b*) differ only in the sign of the second term. These equations may be expanded in the form of a power series in x. With the exception of the term $\pm(\frac{1}{2}x)^2\sin 2\theta$ the series contains only odd powers of x.

The expansion for s is found by dividing by πr^2. In order that the unique character of the second term may be indicated in equations (15*a*) and (15*b*) this term is placed first in the following series:

$$\begin{aligned} s = &\pm\frac{x^2}{\pi}[\tfrac{1}{4}\sin 2\theta] \\ &+\frac{x}{\pi}\left[1-\frac{\log(\sec\theta+\tan\theta)}{\sec\theta\tan\theta}\right] \\ &+\frac{x^3}{\pi}\left[-\frac{1}{3}\cdot\frac{1}{2^3}+\frac{1}{2^3}\cos^2\theta\left\{\frac{1}{2}+\frac{1}{2}\frac{\log(\sec\theta+\tan\theta)}{\sec\theta\tan\theta}\right\}\right] \\ &+\frac{x^5}{\pi}\left[-\frac{1}{5}\cdot\frac{1}{2^2}\cdot\frac{1}{2^5}+\frac{1}{2^2}\cdot\frac{1}{2^5}\cos^2\theta\left(\frac{1}{4}+\frac{3}{4}\cos^2\theta\left\{\frac{1}{2}+\frac{1}{2}\frac{\log(\sec\theta+\tan\theta)}{\sec\theta\tan\theta}\right\}\right)\right] \\ &+\text{...... to infinity} \qquad \text{......(16}a\text{) and (16}b\text{)}. \end{aligned}$$

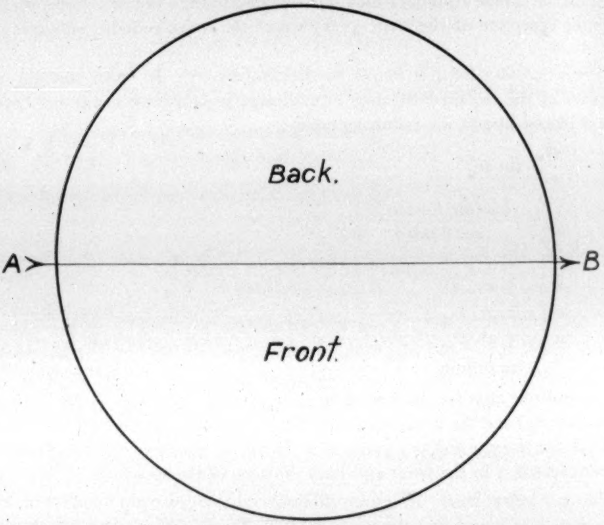
The positive sign for the first term in equation (16) applies to the front of the specimen (16*a*) and the negative sign to the back of the specimen (16*b*). The above series are convergent unless x exceeds 2. It will be shown in the next section that x never exceeds 2 in the front and back portions of the specimen.

Case 3. *Centre Areas.* If figure 1 be modified to correspond to reflection from a centre area, it is found that the locus XSX' does not always terminate on the circumference of the circle. In such cases the limits of integration in equations (5) and (12) are clearly invalid. In other cases equation (5) is valid but not equation (12). In all these cases series (16) is invalid. An alternative series cannot be found except for the case of $\theta = 90°$, to which special consideration will be given. In other cases, the centre areas are relatively unimportant. Indeed for all values of μr exceeding 5 no correction need be applied for them.

§ 3. CHANGE OF ANGLE

It is necessary to define the areas within which equation (16) is valid at different angles. In figures 3 to 8 the areas indicated by the words "front" and "back" both conform to this equation, but the sign of the term $\pi^{-1}(\frac{1}{2}x)^2\sin 2\theta$ is positive for the front area and negative for the back area. The essential conditions for both front and back areas are that the locus XSX' in figure 1 should terminate on the circumference of the circle at points X and X', at which points $q = \pm \tan\theta$. The limits of integration which lead to equation (16) depend on this assumption. In figure 3, which corresponds to $\theta = 0°$, all the loci are parallel to AB and the back is merely a replica of the front. The maximum path is $2r$, so that x does not exceed 2 and there is no central area.

In figures 4 and 5, corresponding to $\theta = 22\frac{1}{2}°$ and $\theta = 30°$ respectively, the boundary of the central area is marked by the loci corresponding to $x = 2$. Inside this area the loci fail to reach the circumference and the conditions of figure 1 are not fulfilled. In figure 5 a small area is shaded. Here the loci reach the circumference but are not terminated at $q = \tan\theta$, so that again equation (16) is inapplicable, although here x is somewhat less than 2.

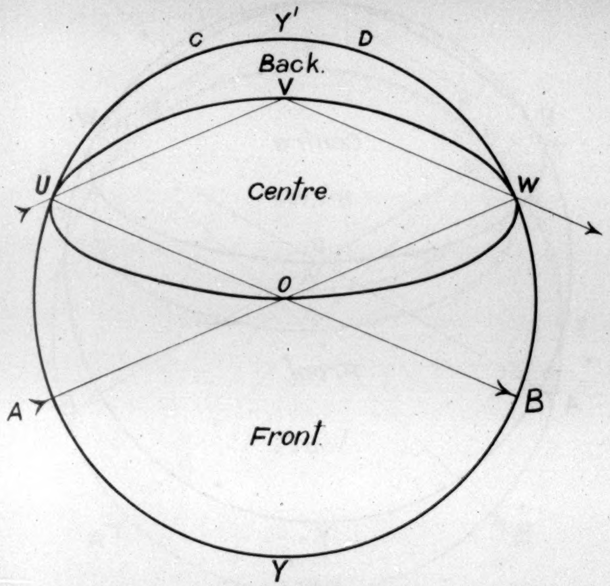


$\theta = 0°$

Figure 3.

In figure 6 ($\theta = 45°$) the boundary of the central area is marked by the locus UOW, corresponding to $x = 2$. The incident ray is tangential at U and the reflected ray is tangential at W. Thus from $\theta = 0°$ to $\theta = 45°$ the boundary of the front area is given by $x = 2$, but beyond 45° this is no longer true. The boundary locus is now displaced towards lower values of x, as may be seen from figure 7 ($\theta = 67\frac{1}{2}°$). The boundary locus is now UTW, which is fixed by the condition that the incident ray is tangential at U and the reflected ray tangential at W. It is clear that x is less than 2 for the locus UTW, but at all higher values of x the incident ray must intersect the specimen before reaching its point of reflection, so that the condition $p=0$, $q=\tan\theta$ is invalidated. At 90°, figure 8, the whole of the specimen is included in the central area, since it is impossible to satisfy the condition $q=\tan\theta$ except on the locus $x=0$.

From $\theta=45^\circ$ to $\theta=90^\circ$ the boundary locus between the front and centre areas is $x=2\sin 2\theta$. To understand what happens if x exceeds this value we may refer



$\theta=22\frac{1}{2}^\circ$

Figure 4.

to figure 6 ($\theta=45^\circ$). A locus such as MNP ends at M and P on the circumference. At M and P the following equation is satisfied:

$$y^2+z^2=1,$$

where y and z have the meaning of equation (6). From the above equation together with equation (6) it can be shown that

$$\{xy\sin\theta+(\tfrac{1}{2}x)^2\}\{q^2\cot^2\theta-1\}=0.$$

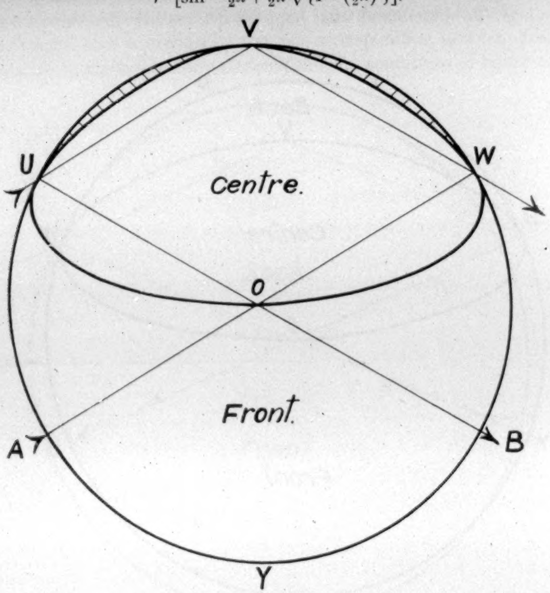
Thus either (*a*) $q=\pm\tan\theta$, the limit of integration in equation (12);

or (*b*) $y=-\frac{1}{4}x\operatorname{cosec}\theta$, whence $q=\pm\sqrt{\{(4/x)^2\sin^2\theta-1\}}$.

Solutions (*a*) and (*b*) are equivalent if $x=2\sin 2\theta$, i.e. at U and W, in either figure 6 or figure 7. It is therefore clear that condition (*a*), which holds at the end of a locus in the front region, is replaced by condition (*b*) on crossing from the front region to the centre region.

It is not possible to deduce an alternative to equation (16) which can be applied to the central area except for $\theta=90^\circ$. Equation (16) for 90° would give

$$r^2\left[\sin^{-1}\tfrac{1}{2}x+\tfrac{1}{2}x\sqrt{\{1-(\tfrac{1}{2}x)^2\}}\right].$$



$\theta=30^\circ$

Figure 5.

Replacing the limit $q=\tan\theta$ at X in figure 1 by the limit appropriate for the central area, $q=\pm\sqrt{\{(4/x)^2\sin^2\theta-1\}}$, we obtain the following value for the area *MNPY* in figure 8:

$$\text{Area } MNPY=r^2\left[2\sin^{-1}\tfrac{1}{4}x+\tfrac{1}{2}x\sqrt{\{1-(\tfrac{1}{4}x)^2\}}\right] \qquad \text{......(17).}$$

This formula replaces equation (16) for the special case in which $\theta=90^\circ$, and may, of course, be expanded in a similar manner.

Table 1. Limiting values of x for front, back and centre areas

θ	Front		Back		Centre	
	Min.	Max.	Min.	Max.	Min.	Max.
0°–30°	0	2	$4\sin\theta$	2	2	$2\sec\theta$
30°–45°	0	2	—	—	2	$2\sec\theta$
45°–90°	0	$2\sin 2\theta$	—	—	$2\sin 2\theta$	$4\sin\theta$

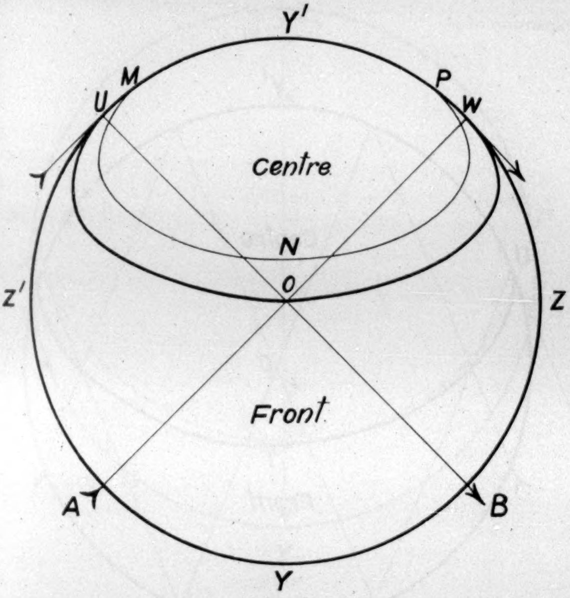
§4. THE ABSORPTION FACTOR: CONTRIBUTION OF FRONT AREAS

(*a*) $\theta<45^\circ$.

Let A_F be the absorption factor calculated for the front area only. A_F

When $$\theta<45^\circ,\quad A_F=\int_{x=0}^{x=2}\frac{ds}{dx}\,e^{-\mu r x}\,dx,$$

where s is given by equation (16*a*).



$\theta=45^\circ$

Figure 6.

Writing equation (16) in the form

$$s=\alpha_1x+\alpha_2x^2+\alpha_3x^3+\ldots,$$ α_r

we have $$\frac{ds}{dx}=\alpha_1+2\alpha_2x+3\alpha_3x^2+\ldots.$$

On integration by parts and suitable rearrangement,

$$A_F=\frac{\alpha_1}{\mu r}(1-e^{-2\mu r})+\frac{2!\,\alpha_2}{\mu^2r^2}\{1-e^{-2\mu r}(1+2\mu r)\}+\frac{3!\,\alpha_3}{\mu^3r^3}\left\{1-e^{-2\mu r}\left(1+2\mu r+\frac{4\mu^2r^2}{2!}\right)\right\}+\ldots\ldots.$$

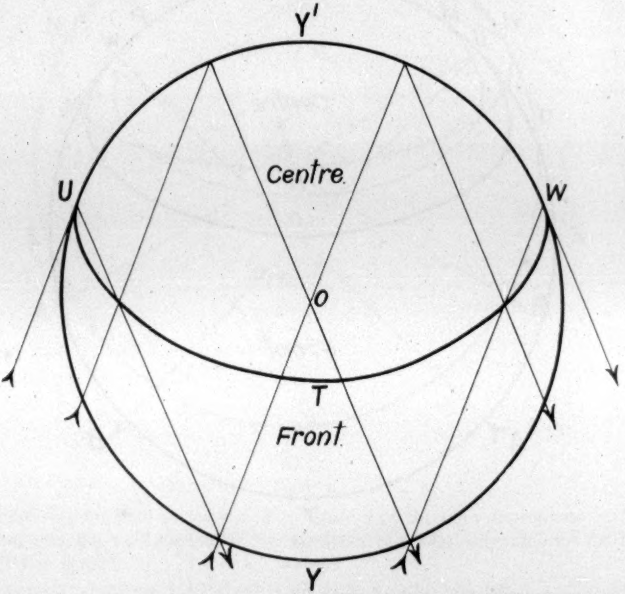
Put $$2\mu r=\rho.$$ ρ

Then

$$A_F=\frac{2\alpha_1}{\rho}(1-e^{-\rho})+\frac{2^2\,2!\,\alpha_2}{\rho^2}\{1-e^{-\rho}(1+\rho)\}+\frac{2^3\,3!\,\alpha_3}{\rho^3}\left\{1-e^{-\rho}\left(1+\rho+\frac{\rho^2}{2!}\right)\right\}$$
$$+\dots+\frac{2^n\,n!\,\alpha_n}{\rho^n}\left(1-e^{-\rho}\sum_{k=0}^{n-1}\frac{\rho^k}{k!}\right)+\dots \qquad \dots\dots(18).$$

$1-e^{-\rho}\sum_{k=0}^{n-1}\frac{\rho^k}{k!}$ may be written $e^{-\rho}\left(e^{\rho}-\sum_{k=0}^{n-1}\frac{\rho^k}{k!}\right)$,

or by expansion of e^{ρ}, $e^{-\rho}\sum_{k=n}^{\infty}\frac{\rho^k}{k!}$.



$\theta=67\frac{1}{2}°$

Figure 7.

It is convenient to write equation (18) thus:

$$\sum_{n=1}^{7}\frac{2^n\,n!\,\alpha_n}{\rho^n}\left(1-e^{-\rho}\sum_{k=0}^{n-1}\frac{\rho^k}{k!}\right)+e^{-\rho}\sum_{n=9}^{\infty}\frac{2^n\,n!\,\alpha_n}{\rho^n}\sum_{k=n}^{\infty}\frac{\rho^k}{k!} \qquad \dots\dots(19).$$

R The order of summation may be reversed in the second or remainder term R, which becomes

$$e^{-\rho}\sum_{l=0}^{\infty}\sum_{n=9}^{\infty}\frac{2^n\,n!\,\alpha_n\,\rho^l}{(n+l)!}.$$

This lies between $e^{-\rho}\sum_{n=9}^{\infty} 2^r\alpha_n$ and $e^{-\rho}\sum_{l=0}^{\infty}\sum_{n=9}^{\infty}\frac{2^n\,9!\,\alpha_n\rho^l}{(9+l)!}$.

A convenient approximation is

$$R=e^{-\rho}\sum_{n=9}^{\infty} 2^n\alpha_n\left\{\frac{1}{2}+\frac{1}{2}\sum_{l=0}^{\infty}\frac{9!}{(9+l)!}\rho^l\right\} \qquad \text{......(20).}$$

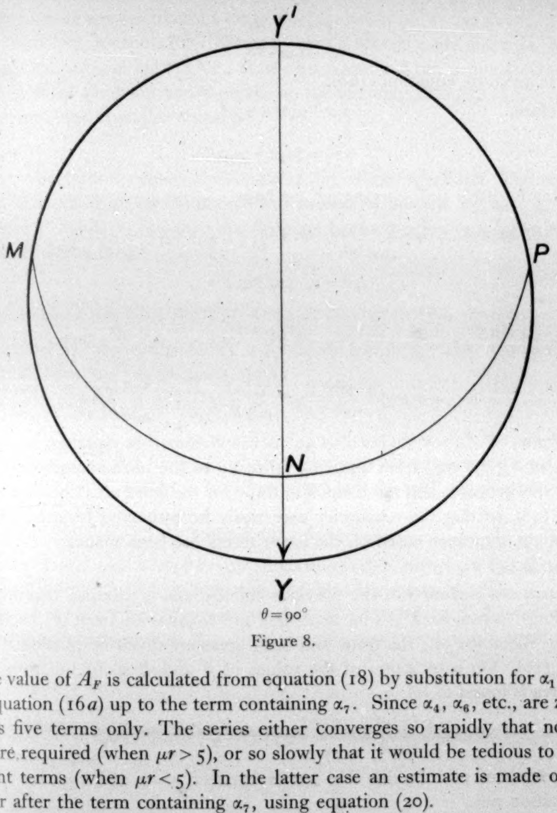


Figure 8.

The value of A_F is calculated from equation (18) by substitution for α_1, α_2, etc., from equation (16*a*) up to the term containing α_7. Since α_4, α_6, etc., are zero, this involves five terms only. The series either converges so rapidly that no further terms are required (when $\mu r > 5$), or so slowly that it would be tedious to calculate sufficient terms (when $\mu r < 5$). In the latter case an estimate is made of the remainder after the term containing α_7, using equation (20).

(*b*) $\theta > 45°$.

When $\theta > 45°$ $$A_F=\int_{x=0}^{x=2\sin 2\theta}\frac{ds}{dx}e^{-\mu r x}\,dx,$$

where s is given by equation (16*a*).

On putting $2\mu r \sin 2\theta = \rho$, we find that equation (18) still holds good but 2^n is replaced by $(2 \sin 2\theta)^n$. The calculations are with this exception the same as for $\theta < 45°$.

§ 5. THE ABSORPTION FACTOR—CORRECTIONS FOR BACK AND CENTRE AREAS

(*a*) *Back areas:* $0 < \theta < 30°$.

A_B

$$A_B = \int_{x=4\sin\theta}^{x=2} \frac{ds}{dx} e^{-\mu r x}\, dx,$$

where s is given by equation (16*b*).

As before,
$$s = \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3$$
and
$$\frac{ds}{dx} = \alpha_1 + 2\alpha_2 x + 3\alpha_3 x^2 + \ldots.$$

Integrating by parts and rearranging, we have

$$A_B = \frac{\alpha_1}{\mu r}(e^{-4\mu r \sin\theta} - e^{-2\mu r}) + \frac{2!\,\alpha_2}{\mu^2 r^2}\{e^{-4\mu r \sin\theta}(1 + 4\mu r \sin\theta) - e^{-2\mu r}(1 + 2\mu r)\} + \ldots.$$

ρ, τ

Put
$$2\mu r = \rho, \quad 4\mu r \sin\theta = \tau.$$

Then
$$A_B = \alpha_1 \left(\frac{4\sin\theta}{\tau} e^{-\tau} - \frac{2}{\rho} e^{-\rho}\right) + 2!\,\alpha_2 \left\{\frac{4^2 \sin^2\theta}{\tau^2} e^{-\tau}(1+\tau) - \frac{2^2}{\rho^2} e^{-\rho}(1+\rho)\right\}$$
$$+ \ldots + \alpha_n n! \left(\frac{4^n \sin^n\theta}{\tau^n} e^{-\tau} \sum_{k=0}^{n-1} \frac{\tau^k}{k!} - \frac{2^n}{\rho^n} e^{-\rho} \sum_{k=0}^{n-1} \frac{\rho^k}{k!}\right) + \ldots \qquad \ldots\ldots(21).$$

The value of A_B for angles 0 to 30° is calculated from equation (21) by substitution for α_1, α_2, etc., from equation (16*b*) up to the term containing α_7. Successive terms get more and more equal to those for the front area obtained by using equation (18), so that the remainder may easily be estimated from equation (20) when a slight allowance based on the lower terms has been made.

(*b*) *Centre areas:* $\theta < 30°$.

The area not included in the previous calculations is grouped together under the heading "centre area". The total area πr^2 consists of front F, back B and centre C. When $\theta < 30°$, the front and back areas are given by putting $x = 2$ into equation (16). On subtraction of the values of F and B so found from πr^2 the central area is found to be

$$C = \frac{4r^2 \cos^2\theta}{\sin\theta} \int_{q=0}^{q=\tan\theta} \sqrt{\{1/(1+q^2) - \cos^2\theta\}}\, dq.$$

The length of path in this area varies between $x = 2$ and $x = 2 \sec\theta$. As an approximation put
$$x = (1 + \sec\theta).$$

A_C

The contribution A_C of the central area is then given by

$$A_C = \frac{4}{\pi} \frac{\cos^2\theta}{\sin\theta} \int_{q=0}^{q=\tan\theta} \sqrt{\{1/(1+q^2)\cos^2\theta\}}\, dq\, e^{-\mu r(1+\sec\theta)} \qquad \ldots\ldots(22).$$

This elliptic integral has already been evaluated in series form, as part of equation (16).

(*c*) *Centre areas*: $30^\circ < \theta < 90^\circ$.

Equation (22) is suitably modified to allow for the difference in area and length of path. As θ approaches 90° this procedure becomes unsatisfactory, owing to the increasing variation in length of path throughout the area. Fortunately it is not necessary to make any calculations for angles just below 90° as the absorption factors may be found by interpolation. When $\theta = 90^\circ$ a special exact formula, which replaces this estimate, is applicable. It is also possible from a modification of the formula used for the front area to make an independent estimate of the correction for the centre area at angles above 45°.

(*d*) $\theta = 90^\circ$.

At 90° the former centre area includes the whole specimen. The absorption factor is deduced from equation (18) but instead of $2\mu r = \rho$, we have $4\mu r = \rho$ and 4^n replaces 2^n, while α_1, α_2, etc., are deduced from equation (17) by an expansion similar to equation (16).

§6. THE ABSORPTION FACTOR: SPECIAL CASES

In general the absorption factor is found by adding together the contributions of the front, back and centre areas, thus:

$$A = A_F + A_B + A_C.$$

A_F is calculated as in equation (18), A_B as in (21) and A_C either as in (22), or by a modification of this formula. A_B is only operative when $0^\circ < \theta < 30^\circ$. For the special case of $\theta = 0^\circ$, equation (21) becomes identical with equation (18), so that we have

$$\text{when } \theta = 0^\circ, \ A = 2A_F.$$

For the special case where $\theta = 90^\circ$, equation (18) is sufficient, but the constants α_1, α_2, etc., are obtained from equation (17) and not equation (16).

When $\mu r > 5$ it is found that except when $\theta = 0^\circ$ and $\theta = 90^\circ$, A_B and A_C may be neglected in comparison with A_F. Further, the remainder term in equation (19) is negligible. In the first term of equation (19) the factor $\left(1 - e^{-\rho} \sum_{k=0}^{n-1} \frac{\rho^k}{k!}\right)$ is approximately equal to unity.

This is not the case for the higher terms. Hence it is important to conclude the series at $n = 7$, thus:

$$\text{when } \mu r > 5 \qquad A = \sum_{n=1}^{7} \frac{\alpha_n \, n!}{\mu^n r^n} \qquad \text{......(23)},$$

where $\alpha_1, \alpha_2 \ldots \alpha_7$ are given as coefficients of $x, x^2 \ldots x^7$ in equation (16*a*).

Equation (23) differs from Claassen's equation (4) because it is a finite series, and therefore the question of convergency does not arise. Equation (23) does not hold for $\theta = 0^\circ$ and $\theta = 90^\circ$.

For extremely large μr only the first term of equation (23) need be considered. From equation (16*a*)

$$\alpha_1 = \frac{1}{\pi}\left\{1 - \frac{\log(\sec\theta + \tan\theta)}{\sec\theta\tan\theta}\right\}.$$

Hence *for extremely large* μr

$$A = \frac{1}{\pi\mu r}\left\{1 - \frac{\log(\sec\theta + \tan\theta)}{\sec\theta\tan\theta}\right\} \qquad \text{......(24).}$$

Equation (24) is virtually equivalent to Rusterholz's formula for the case of great absorption. According to Rusterholz, the expression in brackets is

$$\left\{1 + \frac{\cos^2\theta}{2\sin\theta}\log\frac{\cos 2\theta + \sin\theta}{(1+\sin\theta)(1+2\sin\theta)}\right\}.$$

This can of course be simplified to the form given in equation (24).

When $\mu r > 10$ three terms of equation (23) are sufficient for 1 per cent accuracy.

$$A = \frac{1}{\pi\mu r}\left\{1 - \frac{\log(\sec\theta + \tan\theta)}{\sec\theta\tan\theta}\right\} + \frac{\sin 2\theta}{2\pi\mu^2 r^2} + \frac{1}{\pi\mu^3 r^3}$$
$$\times\left[-\frac{1}{4} + \frac{3}{4}\cos^2\theta\left\{\frac{1}{2} + \frac{1}{2}\frac{\log(\sec\theta + \tan\theta)}{\sec\theta\tan\theta}\right\}\right] \qquad \text{......(25).}$$

At 0° this expression must be doubled, and for angles just above 0° a certain correction is required, to allow for the contribution from the back of the specimen. This contribution shows itself on the film as a doubling of the line and only occurs at extremely low angles. At 90° equation (25) is not applicable, the correct value being, however, only slightly different from that given by this equation. The true formula is as follows:

$$A_{(\theta=90^\circ)} = \frac{1}{\pi\mu r} - \frac{1}{16\pi\mu^3 r^3} \qquad \text{......(26).}$$

The important point to realize is the steady increase in the complexity of the formulae as μr decreases. When $5 < \mu r < 10$ it is necessary to add two terms to equation (25), but the calculations are still not very complicated. When $\mu r < 5$ complications immediately ensue at low angles. It is then necessary to bring in corrections for the higher terms of the series and for the contributions of the centre and back of the specimen. Finally, these corrections become so important that the method of calculation becomes less reliable. When $\mu r < 2$ the Claassen graphical

Table 2. Values of absorption coefficient

μr	$\theta = 0^\circ$		$\theta = 22\frac{1}{2}^\circ$		$\theta = 45^\circ$		$\theta = 67\frac{1}{2}^\circ$		$\theta = 90^\circ$
	Graphical	Calculated	Graphical	Calculated	Graphical	Calculated	Graphical	Calculated	Assumed
5	4·73	4·34	18·40	18·71	48·4	48·0	80·1	81·0	100
4	7·67	7·17	21·30	21·56	51·0	50·8	81·4	82·3	100
3	14·25	13·66	26·75	27·35	56·0	55·3	83·5	84·4	100
2	30·10	30·10	40·50	40·50	64·1	64·1	87·8	88·0	100
1·5	45·60	45·70	53·10	53·10	71·6	71·6	91·0	90·3	100
1	66·80	66·60	71·10	70·60	82·4	81·4	94·3	94·0	100

method supersedes the calculations, and the graphical method can be employed successfully down to $\mu r=0$. Between $\mu r=1$ and $\mu r=5$ it is possible to compare the results obtained by the two methods. It will be seen from table 2 that the agreement is perfect when $\mu r=2$. The calculated values are more correct when $\mu r>2$ and the graphical values when $\mu r<2$.

§ 7. TABLES OF ABSORPTION FACTORS

The two tables of absorption factors were calculated from the formulae given in the present paper except for the range below $\mu r=2$, where Claassen's graphical method was used. The table given in Claassen's paper (table 1, p. 59), which formed the basis of the calculations for this method, was checked and in some instances corrected. The values when $\theta=0°$ especially required emendation. With these changes it has been shown that the graphical method gives exactly the same results as the present formulae for the case where $\mu r=2$ so that there is some reason to suppose that the calculations are accurate for still smaller values of μr, for which the graphical method is even more reliable.

The values when $\mu r<5$ are absolute values of A. When $\mu r>5$ the values are expressed relative to the value for 90°. If it is desired to find the absolute value, it is only necessary to calculate the value of A when $\theta=90°$ from equation (26), which is applicable for all values of μr above 5. The tables are used by plotting the values of A (absolute or relative), as given, against the appropriate value of $\sin^2\theta$. It will be found that the curve obtained is so nearly linear that it can be easily and accurately drawn. For values of θ intermediate between those given in the tables the absorption values can easily be estimated by interpolation. An example is given of the curve for $\mu r=5$.

The values given in these tables are directly applicable to the lines belonging to the layer of zero order in rotation photographs, if the specimen is cylindrical about the axis of rotation. To apply them to powder photographs some modification is required if the specimen is diluted. In place of μr the following substitution is made:

$$\mu r=\frac{\mu}{\rho}\cdot\frac{m}{\pi r h},$$

where μ/ρ is the mass absorption coefficient and m is the mass of powder in a specimen of height h and radius r.

It is assumed that the particles are uniformly distributed in the specimen and that each particle is so small that absorption inside it is negligible. If the diluent has no appreciable absorption coefficient it is only necessary to know μ/ρ, m, r and h. The following technique has been successfully employed by the author. The specimen is diluted with Canada balsam, which also serves as adhesive. It is then mounted on a hair in the form of a cylinder about 0·4 mm. in diameter. After the exposure has been made, the specimen is removed from the powder camera and the average cross-section in the portion exposed to X rays is measured by means of a microscope. It is then placed on a weighed microscope cover slip and the length h is measured, after any irregular portions at the top and bottom have been cut off

Table 3. Absorption factors 100 A when $\mu r < 5$

$\sin^2\theta$	0	0·1464	0·3290	0·5000	0·6710	0·8536	1·00
θ μr	0°	22½°	35°	45°	55°	67½°	90°
0·0	100	100	100	100	100	100	100
0·1	84·70	84·80	—	84·9	—	85·0	85·1
0·2	71·20	71·60	—	71·9	—	72·4	72·9
0·3	60·00	60·60	—	61·4	—	62·7	63·5
0·4	51·00	51·70	—	52·7	—	54·5	55·6
0·5	43·50	44·20	—	45·8	—	47·8	49·0
0·6	36·90	37·80	—	39·8	—	42·3	43·6
0·7	31·40	32·40	—	34·8	—	37·8	39·3
0·8	26·80	27·85	—	30·5	—	33·7	35·6
0·9	23·00	24·10	—	27·1	—	30·5	32·4
1·0	19·77	20·95	—	24·2	—	27·85	29·5
1·1	16·98	18·28	—	21·70	—	25·50	27·15
1·2	14·59	16·00	—	19·54	—	23·50	25·10
1·3	12·56	14·03	—	17·70	—	21·70	23·35
1·4	10·84	12·33	—	16·11	—	20·10	21·80
1·5	9·38	10·91	—	14·69	—	18·66	20·50
1·6	8·11	9·73	—	13·52	—	17·46	19·32
1·7	7·10	8·71	—	12·47	—	16·41	18·24
1·8	6·15	7·80	—	11·54	—	15·42	17·30
1·9	5·37	7·02	—	10·74	—	14·59	16·44
2·0	4·71	6·35	—	10·05	—	13·84	15·67
2·1	4·16	5·79	—	9·44	—	13·15	14·93
2·2	3·67	5·31	—	8·89	—	12·50	14·26
2·3	3·24	4·86	—	8·38	—	11·89	13·65
2·4	2·865	4·47	—	7·91	—	11·35	13·09
2·5	2·55	4·12	—	7·50	—	10·86	12·56
2·6	2·27	3·82	—	7·11	—	10·40	12·11
2·7	2·02	3·55	—	6·75	—	9·98	11·67
2·8	1·803	3·30	—	6·41	—	9·62	11·27
2·9	1·607	3·08	—	6·10	—	9·25	10·89
3·0	1·436	2·885	4·43	5·82	7·24	8·89	10·54
3·1	1·288	2·705	4·22	5·58	6·97	8·57	10·21
3·2	1·159	2·55	4·02	5·35	6·72	8·30	9·90
3·3	1·049	2·415	3·84	5·14	6·48	8·04	9·61
3·4	0·955	2·29	3·68	4·95	6·26	7·78	9·33
3·5	0·871	2·17	3·525	4·77	6·05	7·55	9·06
3·6	0·796	2·06	3·385	4·60	5·85	7·33	8·81
3·7	0·729	1·968	3·255	4·44	5·66	7·11	8·58
3·8	0·670	1·875	3·14	4·29	5·49	6·92	8·36
3·9	0·617	1·787	3·03	4·15	5·33	6·73	8·15
4·0	0·568	1·706	2·925	4·02	5·17	6·53	7·94
4·1	0·525	1·629	2·825	3·89	5·02	6·35	7·74
4·2	0·488	1·563	2·73	3·77	4·88	6·18	7·55
4·3	0·453	1·500	2·645	3·66	4·75	6·02	7·38
4·4	0·420	1·445	2·555	3·56	4·62	5·87	7·21
4·5	0·391	1·390	2·485	3·47	4·50	5·73	7·05
4·6	0·364	1·343	2·41	3·38	4·39	5·61	6·89
4·7	0·340	1·300	2·34	3·29	4·28	5·47	6·75
4·8	0·316	1·259	2·275	3·21	4·18	5·35	6·61
4·9	0·2945	1·222	2·21	3·13	4·08	5·25	6·47
5·0	0·2755	1·189	2·155	3·05	3·99	5·14	6·35

Table 4. Relative values of A for large values of μr

$\sin^2\theta$	0	0·1464	0·3290	0·5000	0·6710	0·8536	1·00
θ / μr	0°	$22\frac{1}{2}$°	35°	45°	55°	$67\frac{1}{2}$°	90°
5·0	4·34	18·71	33·93	48·00	62·8	81·0	100
5·5	3·50	17·78	32·95	47·05	62·0	80·4	100
6·0	2·91	17·01	32·10	46·25	61·3	79·9	100
6·5	2·44	16·40	31·40	45·60	60·7	79·5	100
7·0	2·12	15·89	30·80	45·05	60·2	79·2	100
7·5	1·83	15·46	30·30	44·50	59·8	78·9	100
8·0	1·61	15·09	29·85	44·05	59·4	78·6	100
9·0	1·26	14·45	29·10	43·35	58·7	78·1	100
10·0	1·02	13·98	28·55	42·75	58·2	77·7	100
11·0	0·84	13·59	28·10	42·25	57·8	77·4	100
12·0	0·69	13·26	27·70	41·85	57·4	77·2	100
13·0	0·59	13·00	27·40	41·50	57·1	77·0	100
14·0	0·50	12·78	27·15	41·25	56·9	76·8	100
15·0	0·44	12·59	26·90	41·00	56·7	76·7	100
20·0	0·25	11·93	26·05	40·20	55·9	76·2	100
25·0	0·16	11·55	25·60	39·70	55·4	75·8	100
30·0	0·11	11·29	25·25	39·35	55·1	75·6	100
40·0	0·06	10·97	24·85	38·95	54·8	75·3	100
50·0	0·04	10·79	24·60	38·70	54·5	75·1	100
75·0	0·02	10·54	24·30	38·40	54·2	74·9	100
100	0·01	10·42	24·10	38·20	54·0	74·8	100
200	0·00	10·18	23·90	38·00	53·8	74·6	100
∞	0·00	10·07	23·65	37·70	53·6	74·4	100

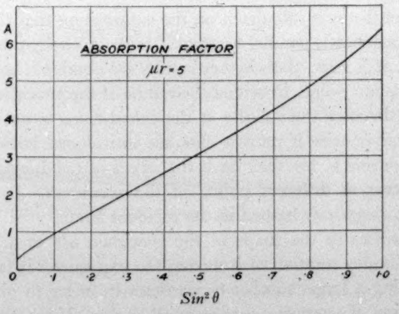


Figure 9.

with a razor blade. The Canada balsam is dissolved away by means of toluene, which is afterwards soaked up with blotting paper. The hair can then be removed and finally only the powder is left on the cover slip, which is reweighed, the weight of powder being obtained by difference. If necessary, where the absorption coefficient of the powder is small a correction may be made for absorption in the Canada balsam. This proceeds as for a powder containing ingredients of different absorption coefficients.

Suppose that the powder contains two ingredients, a mass xm of an ingredient with mass absorption coefficient μ_1/ρ_1 and a mass $(1-x)\,m$ of a second ingredient with mass absorption coefficient μ_2/ρ_2. The value of μr is then given by

$$\mu r=\left[\left(\frac{\mu_1}{\rho_1}\,xm\right)+\left\{\frac{\mu_2}{\rho_2}\,(1-x)\,m\right\}\right]\frac{1}{\pi rh}.$$

Evidently only the value of x is required in addition to the previous data. It must be emphasized that the particles of the two ingredients need to be perfectly mixed, so that no portion of the specimen contains enough of one material to cause an abnormal absorption.

Finally, a word of warning is necessary with regard to the utility of the absorption factor. The observed intensities are attributed to a number of factors. Some depend on the structure of the crystals while others depend on the angle of reflection, changing steadily as the angle passes from small to large values. Amongst the factors concerned must be included the temperature factor, which causes a steady decrease in the intensities as θ increases. This is the opposite of the absorption factor, which reduces the intensities at low angles much more than those at high angles. In many instances the effects of the absorption factor and the temperature factor almost cancel one another. Consequently a comparison of observed intensities with calculated intensities will often show better agreement if no allowance is made for the absorption correction, unless at the same time the values are corrected for the temperature effect.

The absorption factor is calculated on the assumption that the whole specimen is bathed in a homogeneous parallel beam of X rays. Actually there is no such thing as a parallel beam of X rays. Satisfactory results are obtained, however, if the beam diverges from a distant point. Provided the radius of the specimen is small in comparison with this distance the validity of the calculations is not seriously affected. Trouble will certainly arise if narrow slits are introduced between the source of X rays and the specimen, for they have the effect of producing variations in the intensity-distribution at different points of the specimen. In other words, the specimen must be *completely* bathed in the incident beam.

Slight irregularities in the shape of the specimen are to a large extent compensated by continuous rotation of it during the exposure. It is essential to use a specimen containing a large number of particles in order to obtain a sufficiently uniform distribution of scattering matter at all angles. If the powder is packed in Lindemann glass tubing, this must be so thin-walled that absorption in the glass is negligible.

§ 8. ACKNOWLEDGMENT

The author thanks Professor W. L. Bragg, F.R.S. for his kind interest in the work, which was carried out in the Physical Laboratories of the University of Manchester.

REFERENCES

(1) L. W. McKeehan. *J. Frankl. Inst.* **193**, 231 (1922).
(2) G. Greenwood. *Phil. Mag.* (7), **3**, 963 (1927).
(3) H. Möller and A. Reis. *Z. phys. Chem. A.* **139**, 425 (1928).
(4) A. Claassen. *Phil. Mag.* (7), **9**, 57 (1930).
(5) A. Rusterholz. *Helv. Physica Acta*, **4**, 68 (1931).

538.313

THE MAGNETIC FORCE AT A POINT ON ITS AXIS DUE TO A CURRENT IN A HELICAL COIL OF ONE TURN

By G. F. C. SEARLE, Sc.D., F.R.S.,
University Lecturer in Experimental Physics, Cambridge

Received June 1, 1935

ABSTRACT. The incomplete circuit formed by a single turn *ABC* of a helix of pitch *p*, which lies on a cylinder of radius *a*, is completed by a straight link *CA*, of length *p*, parallel to the axis of the cylinder. Through any point *O* on this axis are taken rectangular axes *Ox*, *Oy*, *Oz*. Of these, *Ox* is the axis of the cylinder, and *Oz* is parallel to the shortest distance between *Ox* and *CA*. The components H_1, H_2, H_3 of the magnetic force at *O*, due to a current in the circuit, are found for the case when p/a is small.

---

§ 1. MAGNETIC FORCE AT ORIGIN

Let the helix *ABPC*, figure 1, lie on a cylinder of radius *a*; let it be right-handed, with pitch *p*. The circuit is completed by a portion *CDA*, of length *p*, of a generator of the cylinder. The axis of the helix is the axis of *x*, and right-handed axes are used. We will take as origin *O* the point on the axis of *x* at which the magnetic force is to be found. The mean plane of the helix, parallel to *Oyz*, cuts the helix in *B* and the axis in *N*, where $x=h$, and bisects *CDA* in *D*. The point *P* lies on the helix, the plane *POx* has turned in the positive direction through angle θ from the plane *zOx*, and $OP=r$. The link *CDA* is in the plane *zOx*. If the coordinates of *P* be ξ, η, ζ, and if $p/2\pi=q$, we have

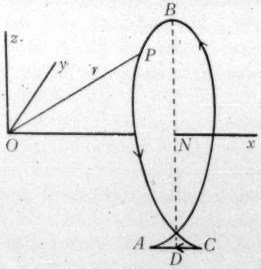


Figure 1.

$$\xi=h+p\theta/2\pi=h+q\theta, \quad \eta=-a\sin\theta, \quad \zeta=a\cos\theta.$$

Then
$$d\xi/d\theta=q, \quad d\eta/d\theta=-a\cos\theta, \quad d\zeta/d\theta=-a\sin\theta.$$

If ds be an element of arc, $ds/d\theta = \sqrt{(q^2+a^2)} = b$. If λ, μ, ν be the direction cosines of the forward direction of ds,

$$\lambda = q/b, \quad \mu = -a \cos\theta/b, \quad \nu = -a \sin\theta/b.$$

Let dH_1, dH_2, dH_3 be the components, along Ox, Oy, Oz, of the magnetic force at O due to the element ds, or $bd\theta$, at P. Then

$$\begin{aligned} dH_1 &= ibd\theta r^{-3}(\nu\eta - \mu\zeta) = ia^2 d\theta r^{-3}, \\ dH_2 &= ibd\theta r^{-3}(\lambda\zeta - \nu\xi) = iad\theta r^{-3}\{h\sin\theta + q(\cos\theta + \theta\sin\theta)\}, \\ dH_3 &= ibd\theta r^{-3}(\mu\xi - \lambda\eta) = iad\theta r^{-3}\{-h\cos\theta + q(\sin\theta - \theta\cos\theta)\}. \end{aligned}$$

We have $$r^2 = a^2 + (h+q\theta)^2.$$

If $q\theta$ be small compared with l, where $l^2 = a^2 + h^2$, we have

$$r^{-3} = l^{-3}\{1 + c_1 q\theta + c_2 q^2\theta^2 + c_3 q^3\theta^3 + \dots\},$$

where $$c_1 = -3h/l^2, \quad c_2 = -\tfrac{3}{2}(a^2 - 4h^2)/l^4, \quad c_3 = \tfrac{5}{2}h(3a^2 - 4h^2)/l^6.$$

For H_1 we have, as far as q^3,

$$H_1 = ia^2 \int_{-\pi}^{\pi} r^{-3} d\theta = \frac{ia^2}{l^3}\int_{-\pi}^{\pi}(1 + c_1 q\theta + \dots)\, d\theta = \frac{2\pi i a^2}{l^3}(1 + \tfrac{1}{3}c_2\pi^2 q^2)$$

$$= \frac{2\pi i a^2}{l^3}\left\{1 - \frac{\pi^2 q^2 (a^2 - 4h^2)}{2l^4}\right\} \qquad \text{......(1).}$$

The axial component H_1 of the magnetic force equals, accurately, that due to a uniform cylindrical current sheet, of radius a and strength i/p, extending from the plane $x = h - \frac{1}{2}p$ to the plane $x = h + \frac{1}{2}p$. Accurately

$$H_1 = \frac{2\pi i}{p}\left\{\frac{h + \frac{1}{2}p}{\{(h+\frac{1}{2}p)^2 + a^2\}^{\frac{1}{2}}} - \frac{h - \frac{1}{2}p}{\{(h-\frac{1}{2}p)^2 + a^2\}^{\frac{1}{2}}}\right\}.$$

When we expand as far as p^2, we obtain (1).

We next consider H_2. As far as q^3,

$$\{h\sin\theta + q(\cos\theta + \theta\sin\theta)\}\, r^{-3} = \{f_0 + f_1 q + f_2 q^2 + f_3 q^3\}\, l^{-3},$$

where $f_0 = h\sin\theta$, $\quad f_1 = \cos\theta + \theta\sin\theta\,(1 + c_1 h)$,

$f_2 = c_1\theta\cos\theta + \theta^2\sin\theta\,(c_1 + c_2 h)$, $\quad f_3 = c_2\theta^2\cos\theta + \theta^3\sin\theta\,(c_2 + c_3 h)$.

When the following integrations extend from $-\pi$ to π, we have

$$\begin{array}{llll} \int \sin\theta d\theta = 0, & \int \cos\theta d\theta = 0, & \int \theta\sin\theta d\theta = 2\pi, & \int \theta\cos\theta d\theta = 0, \\ \int \theta^2\sin\theta d\theta = 0, & \int \theta^2\cos\theta d\theta = -4\pi, & \int \theta^3\sin\theta d\theta = 2\pi^3 - 12\pi, & \int \theta^3\cos\theta d\theta = 0. \end{array}$$

If H_2' be the part of H_2 which is due to the helix ABC,

$$H_2' = \frac{ia}{l^3}\int_{-\pi}^{\pi}(f_0 + f_1 q + \dots)\, d\theta$$

$$= \frac{2\pi i a q}{l^3}\left[1 - \frac{3h^2}{l^2} - \frac{q^2}{2l^2}\left\{3(\pi^2 - 8)\frac{a^2 - 4h^2}{l^2} - 5(\pi^2 - 6)\frac{h^2(3a^2 - 4h^2)}{l^4}\right\}\right].$$

The current along CA from C to A produces at O magnetic force H_2'' along Oy. We have

$$H_2''=ia\int_{-\frac{1}{2}p}^{\frac{1}{2}p}\frac{dt}{\{a^2+(h+t)^2\}^{\frac{3}{2}}}=\frac{ia}{l^3}\int_{-\pi q}^{\pi q}(1+c_1t+c_2t^2+\ldots)\,dt$$

$$=\frac{2\pi iaq}{l^3}\left\{1-\frac{\pi^2q^2\,(a^2-4h^2)}{2l^4}\right\}.$$

If H_2 be the total magnetic force at O along Oy, $H_2=H_2'+H_2''$, and

$$H_2=\frac{2\pi iaq}{l^3}\left\{\frac{2a^2-h^2}{l^2}-\frac{(\pi^2-6)\,q^2}{2l^6}(4a^4-27a^2h^2+4h^4)\right\} \qquad \text{......(2)}.$$

We now find H_3. As far as q^3, we have

$$\{-h\cos\theta+q\,(\sin\theta-\theta\cos\theta)\}\,r^{-3}=\{g_0+g_1q+g_2q^2+g_3q^3\}\,l^{-3},$$

where $g_0=-h\cos\theta$, $\quad g_1=\sin\theta-\theta\cos\theta\,(1+c_1h)$,

$$g_2=c_1\,\theta\sin\theta-\theta^2\cos\theta\,(c_1+c_2h), \quad g_3=c_2\theta^2\sin\theta-\theta^3\cos\theta\,(c_2+c_3h).$$

On integration from $-\pi$ to π, we have

$$\int g_0d\theta=0, \quad \int g_1d\theta=0, \quad \int g_3d\theta=0$$

$$\int g_2d\theta=2\pi\,(3c_1+2c_2h)=-6\pi h\,(4a^2-h^2)\,l^{-4}.$$

Hence
$$H_3=\frac{ia}{l^3}\int_{-\pi}^{\pi}(g_0+g_1q+\ldots)\,d\theta$$

$$=-\frac{6\pi iaq^2h}{l^7}(4a^2-h^2) \qquad \text{......(3)}.$$

§ 2. MAGNETIC FORCE NEAR ORIGIN

If we wish to find the components of the magnetic force at a point Q, of coordinates x, y, z, near O, we must, in place of the value of dH_1 used in § 1, write

$$dH_1=ibd\theta\,r_Q^{-3}\{\nu\,(\eta-y)-\mu\,(\zeta-z)\},$$

where $\quad r_Q^2=(\xi-x)^2+\ldots=r^2\{1-2\,(\xi x+\eta y+\zeta z)/r^2+(\xi x^2+\ldots)/r^2\}.$

Similar expressions hold for dH_2, dH_3. We first expand r_Q^{-3} in powers of x, y, z and of r^{-1} and then expand r^{-3}, r^{-5}, ... in powers of $q\theta$. We then insert the values of ξ, η, ζ, λ, μ, ν in terms of θ and integrate with respect to θ as in § 1.

§ 3. A PAIR OF HELICES

If the helix be turned about its axis through 180°, the signs of H_2 and H_3 are reversed. Hence we obtain the following accurate result:

Let equal currents flow in the same direction in two equal coaxial helices ABC, $A'B'C'$, each of one turn, with their mean planes coincident. If the angle between the two planes containing (1) the axis and the link AC, (2) the axis and the link $A'C'$ be π, the magnetic force at any point on the axis has no component at right angles to the axis.

§ 4. COMPUTATION BY MAGNETIC SHELL

It is, perhaps, surprising that the leading term in H_2 should be of the *first* order in q. We can, however, verify the result. The helix ABC (figure 1) and the link CA form the edge of a magnetic shell of strength i. We take the shell to have a cylindrical part and a plane part. The cylindrical part has Ox for axis and extends from the helix in the positive direction of x. The plane part, which closes the positive end of the cylindrical part, is a disk of radius a with Ox for axis. The part of the shell on the positive side of the plane $x=h+\frac{1}{2}p$ is symmetrical about Ox, and the magnetic force at O due to it has no component along Oy. The remainder of the shell is a strip of width k, where

$$k=\tfrac{1}{2}p\,(1-\theta/\pi)=q\,(\pi-\theta).$$

With the direction of the current employed in § 1, the magnetism on the outer side of the shell is positive. If da be the thickness of the shell and if σ be the surface density of magnetism on the outer surface, $\sigma da=i$.

The magnetic force at O, due to a pole m at a point ξ, η, ζ on the cylinder, has a y-component

$$-\frac{m\eta}{(a^2+\xi^2)^{\frac{3}{2}}}=\frac{ma\sin\theta}{r^3}.$$

If a be increased to $a+da$, the y-component due to m becomes

$$m\sin\theta\left\{\frac{a}{r^3}+\frac{d}{da}\left(\frac{a}{r^3}\right)da\right\}.$$

Hence, if dH_2 be the y-component of the magnetic force at O due to an element of shell of area dS at ξ, η, ζ, we have

$$dH_2=\sigma dS\sin\theta\,\frac{d}{da}\left(\frac{a}{r^3}\right)da=idS\sin\theta\,\frac{\xi^2-2a^2}{r^5}.$$

If we do not go beyond the first power of q, we may put $\xi^2=h^2$, $r^2=l^2=a^2+h^2$, and $dS=kad\theta$. Then, to the first power of q,

$$H_2=\frac{iaq\,(h^2-2a^2)}{l^5}\int_{-\pi}^{\pi}(\pi-\theta)\sin\theta d\theta=\frac{2\pi iaq\,(2a^2-h^2)}{l^5},$$

as found in § 1.

§ 5. ACKNOWLEDGMENT

Mr K. T. Li, Emmanuel College, Cambridge, B.Sc., National Central University, Nankin, China, has kindly verified the analysis in this paper.

537.322.2:546.57

LONGITUDINAL THERMOELECTRIC EFFECT: (5) SILVER

By J. L. CH'EN, M.S. (Yenching)

AND

WILLIAM BAND, M.Sc., Professor of Physics,
Yenching University, Peiping, China

Received April 10, 1935

ABSTRACT. Improvements in technique are described by means of which the experimental results become more consistent than before, and the analysis becomes simpler to carry out. The homogeneous thermoelectric constants for silver are found, and their dependence upon longitudinal tension is shown. A critical temperature near 200° C. for zero tension is found at which the thermoelectric constants suffer an abrupt change. Reasons are given for supposing that there is an intimate connection between this transition and the ordinary elastic limit of the wire.

§ 1. IMPROVEMENTS IN TECHNIQUE

THE general method and idea of this work can be learned from the previous papers under the present general title (1, 2, 3). The thermoelectric e.m.f. in this case was measured with the Paschen galvanometer previously described in part 3, but the thermal regulator was similar to that described in part 2. The chief improvement over the method used in part 3 was to surround the heating-system with a cylindrical metal jacket containing a water-cooled cavity: this cylinder was closed at both ends except for small apertures at the axes for the specimen and thermocouples, and it maintained the outside conditions for the heating system at a definite known temperature throughout.

To improve the probe method of investigating the temperature-distribution, the probe elements were led into the axial hole from opposite ends of the axis and joined in the middle to form the thermojunction. Although the two elements have not quite the same thermal properties, yet this device, when made to slide to and fro along the axis in exploring the temperature-distribution, certainly creates much less disturbance of the temperatures than did the previous probe entering entirely from one end. As has previously been pointed out, the probe, while for obvious reasons not as precise as could be wished, is the only practicable means available for the purpose in view. The mechanical and thermal maltreatment of the specimen involved in the welding of a thermo-junction on to or into it, most decidedly has to be avoided if any significance is to be attached to the e.m.fs. as due to a homogeneous metal. However, to check errors arising in the positions of the temperatures given by the travelling junction, it is necessary to have one or more junctions fixed in position along the axis; these also suffer the same temperature error due to their superficial contact with the specimen, but it is assumed that the axial hole has an almost uniform temperature-distribution in planes perpendicular to the axis.

These fixed thermocouples also served to check the constancy of the temperatures at fixed points while the travelling couple was moving through the system during the mapping process.

In the previous work the analysis was carried out by trial and error; in the present work a graphical device eliminates the necessity for this. The integrals of the temperature-distribution in the equation:

$$E = a \int G . dT + b \int G^2 . dT$$

a, b

are found as before for each temperature-distribution, and then the straight lines are plotted against the axes *a* and *b* as the two variables. The common point of intersection of the whole set of lines gives, theoretically, the required values of *a* and *b*. Actually, the weighted centre of the smallest area containing all these lines will in practice be the point accepted. A visual picture of the precision of the result is thus very nicely obtained.

§ 2. RESULTS

In figure 1 are shown seven typical temperature-distributions split into straight-line sections for analysis. For each temperature-distribution the specimen was carried through the set of tensions investigated, and the resulting e.m.fs. were recorded. It was early noted that after a certain tension had been passed the changes in e.m.f. ceased to be reversible and a permanent alteration of the {e.m.f., tension} curves took place. The wire was therefore replaced and the observations were repeated, care being taken always to keep the tension below this critical value.

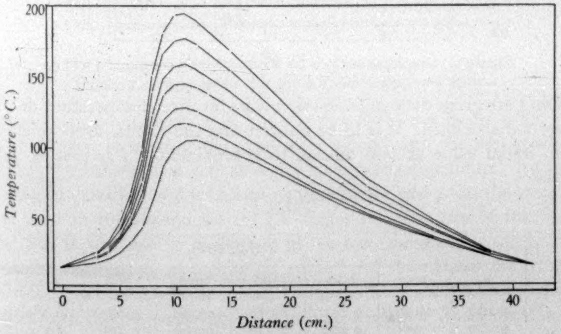


Figure 1. Typical temperature-maps.

In figure 2 is shown the graphical analysis of the data for zero tension on the first specimen. The common point of intersection of all the lines is well defined, indicating a satisfactory precision for the resulting values of the thermoelectric constants. The values indicated are

$$a = -6{\cdot}1 \times 10^{-10},$$

and

$$b = 8{\cdot}6 \times 10^{-12}.$$

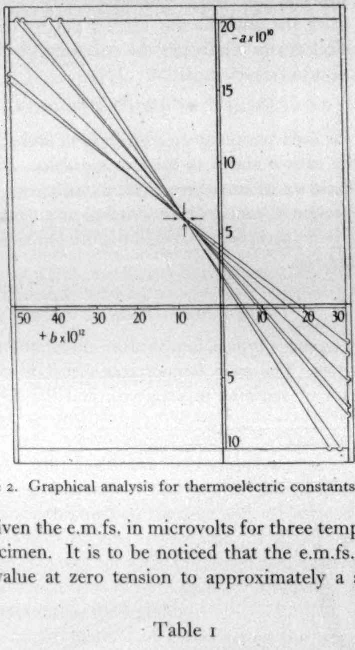


Figure 2. Graphical analysis for thermoelectric constants *a* and *b*.

In table 1 are given the e.m.fs. in microvolts for three temperature-distributions for the second specimen. It is to be noticed that the e.m.fs. drop quickly from a maximum initial value at zero tension to approximately a saturation value near 300 kg./cm.²

Table 1

Curve	S (kg./cm.²)							
	0	100	200	300	400	500	600	700
1	0·166	0·129	0·105	0·080	0·080	0·080	0·080	0·080
2	0·270	0·221	0·180	0·173	0·169	0·166	0·170	0·172
3	0·295	0·229	0·230	0·225	0·241	0·252	0·247	0·252

The values obtained for zero tension are, in this case

$$a = -\ 6{\cdot}2 \,.\, 10^{-10}\ \text{V.-cm.}/(^\circ\ \text{C.})^2$$

$$b = \ 14{\cdot}5 \,.\, 10^{-12}\ \text{V.-cm.}^2/(^\circ\ \text{C.})^3$$

In figure 3 are given the values of the constants as functions of tension. The curves are drawn from the empirical formulae:

$$-a\,.\,10^{10}=3{\cdot}4+2{\cdot}8\times\exp\,(-0{\cdot}00625\,.\,S),$$
$$b\,.\,10^{12}=5{\cdot}0+9{\cdot}5\times\exp\,(-0{\cdot}00369\,.\,S).$$

The asymptotes, unlike those found in the case of copper(1), appear to be practically parallel with the S axis as in normal saturation curves.

It will be noticed that the two specimens gave almost identical values for the constant a but that the two values for b are quite different. It was observed that the general behaviour of the e.m.f. with changing tension was the same for the two specimens. Since both specimens were taken from the same piece of wire it is not possible to explain the difference in b as due to a difference in the metal or its impurities. However, in view of the experience gained with the first specimen, the second specimen was not heated to so high a maximum temperature. It may be that there is a rapid variation of b with temperature as we pass from the range of temperatures used in the second set to the higher temperatures of the first set of readings. But in that case it is difficult to see how the analysis can give definite results on the assumption of a constant value for b throughout the temperatures existing in the specimen (from about 20° C. up to the maximum used, say 150° C.). Slight twisting of the wire may account for the change in b, which must in that case be very sensitive to the twist, for care was taken not to twist the wire during mounting in the apparatus.

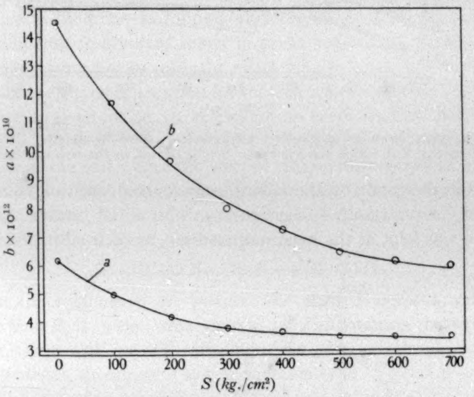


Figure 3. Dependence of a and b upon longitudinal strain.

These effects require further detailed investigation of the dependence of the e.m.f. upon temperature-range, and upon state of torsion of the material.

The afore-mentioned transition was next investigated, use being made of several different specimens in turn as each was carried over the transition temperature.

The specimen was heated to a temperature as near under the transition as possible without actually passing it, the normality being checked by carrying the specimen through a tension cycle. The process was repeated at a slightly higher maximum temperature (at the peak of the distribution curve) and an entirely different {tension, e.m.f.} curve resulted; see figure 4. The initial heating was carried out, while the wire was kept at the maximum tension to be used in the cycle. It was found, first, that if this maximum tension was increased, the maximum temperature required to produce the transition was not so high. If the wire was stretched beyond its elastic limit at room-temperature its behaviour corresponded with the phenomena found after the change at higher temperatures and smaller tension. Figure 5 shows how the transition temperature depended upon the maximum

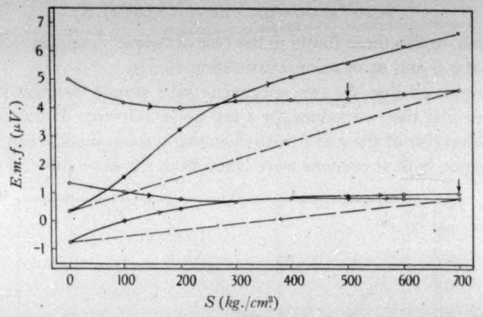


Figure 4. Discontinuity produced at elastic limit in {e.m.f., tension} curves. The arrows indicate the tensions S at which the hysteresis first appeared in the two specimens.

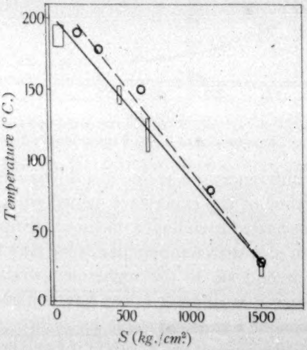


Figure 5. Thermoelectric discontinuity and the elastic limit. The full line and rectangles show the thermoelectric data, the dotted line and circles the mechanical data.

tension. It seemed most probable that this curve followed the course of ordinary elastic limit as it changes with temperature. A direct test of this suggestion gave the dotted curve in the same figure. Evidently the correct explanation is that the part of the wire heated above the critical temperature is stretched beyond its elastic limit and so behaves abnormally, either in that it forms a thermocouple with the normal wire, or in that its stretching allows bodily motion of the rest of the wire through the temperature-field.

§ 3. NOTE ON CONVENTIONS AS TO SIGNS

In the *International Critical Tables* the sign convention for the Thomson coefficient is that when the electrons have positive specific heat the effect is positive. This convention is modified as follows, in the present work, to include the new e.m.f.: If a temperature-gradient tends to cause a spontaneous flow of electrons down the gradient from hot to cold, the effect is to be taken as positive. This convention includes the above convention for the Thomson coefficient, and also enables a definite sign to be given to the two homogeneous coefficients studied in the present work. Thus if, as in the effect for silver, the spontaneous positive current flows out of the specimen at the end where the steeper temperature-gradient terminates, the electrons must be flowing up the steeper gradient and down the flatter gradient; the effect is then negative. With this convention, the formula used for analysis in previous papers on this subject, has been written

$$\mathbf{F} = t\mathbf{G} + a\,(\mathbf{G}\,.\,\mathbf{G})^{\frac{1}{2}}\,\mathbf{G} + b\,(\mathbf{G}\,.\,\mathbf{G})\,\mathbf{G},$$

because when the coefficients are positive the positive current, supposed to be driven by the field **F**, is in the same sense as the temperature-gradient **G**, the temperature increasing with positive displacement. The electrons flow down the temperature-gradient, or opposite to the direction of **G**.

REFERENCES

(1) P. C. Feng and W. Band. *Phys. Soc. Proc.* **46**, 515 (1934).
(2) T. H. Pi and W. Band. Page 852 of this volume.
(3) M. K. Li and W. Band. Page 859 of this volume.

538.66:546.74

THERMOMAGNETIC HYSTERESIS IN NICKEL WIRE

By Y. K. HSÜ, M.S. (Yenching)

AND

WILLIAM BAND, M.Sc., Professor of Physics,
Yenching University, Peiping, China

Received April 10, 1935

ABSTRACT. The thermomagnetic e.m.f. in a nickel wire of diameter 1·0 mm. is produced by longitudinal magnetic fields and by longitudinal stretching in one-half of each of a set of various temperature-distributions. The hysteresis is examined as dependent both upon magnetism and upon tension. It is found that for zero tension the results agree with those of Broili⁽¹⁾ to a first approximation, although there is still a measurable increase in e.m.f. not detected by Broili when the maximum temperature is raised above the Curie point. This is interpreted as due to the so-called Benedick e.m.f. depending on the temperature-gradient. The Curie-point increase is still greater in the results for stretched wire. The form of the hysteresis curve also becomes anomalous in that the e.m.f. reverses sign, relative to the original value, in two quadrants of the loop. The change in Thomson potential-gradient produced by various magnetizing fields, at different tensions, is shown graphically as a function of temperature, to a first approximation obtained by neglecting the Benedick e.m.f.

§ 1. INTRODUCTION

The thermomagnetic properties of nickel wire have been studied by Heinz Broili⁽¹⁾. The specimen was centrally heated to a uniform temperature while the ends were maintained at a constant lower temperature. The part of the wire in the temperature-drop at one end was subjected to a magnetic field, while the part of the wire in the other temperature-drop was outside the magnetic field. The electromotive force produced in the wire by this arrangement was studied and found to show positive hysteresis. The e.m.f. was also found to vary with temperature-difference, but when the temperature-gradient was changed without change of the temperature-difference, no detectable change in the e.m.f. occurred.

Work of somewhat similar nature was done by Chang and Band⁽²⁾ on iron wire. Here the hysteresis was found to be negative, and a qualitative theory was proposed to explain this result. This particular theory, however, would demand a negative hysteresis even in nickel, contrary to Broili's results. More work was therefore planned on nickel wire to check this point.

In the work of Pi and Band⁽³⁾ nickel was first examined for the Benedick effect, which according to Broili's results either did not exist or was masked by the larger modification of the Thomson e.m.f. in the magnetic field. The work gave positive results; the e.m.f. was however of a smaller order than that observed by Broili, the maximum change produced by a magnetizing field of about 200 gauss being only a few microvolts. In this work all the e.m.fs. were considered as being due respectively to the magnetic modification of the Thomson potential-gradient and of the two new thermoelectric constants *a* and *b*, which were associated with the square and the cube of the temperature-gradient$^{(4)}$.

In the work reported in the present paper, nickel* has again been examined in an arrangement essentially similar to that of Broili. A uniform magnetic field was superposed on a part of the wire also suffering a temperature-gradient, while the parts of the wire which had varying magnetization were kept approximately at a uniform temperature. In addition to this, the wire could be stretched on one side to a different tension from that on the other, by means of a mechanical clamp electrically insulated from the specimen. Thus the pure thermomechanical e.m.f. caused by the modification of Thomson potential-gradient under longitudinal strain could be investigated. Also, as a further improvement over Broili's work, not only could the temperature-difference be obtained by a fixed thermocouple, but the whole temperature-distribution could be mapped throughout the system by means of thermocouple probes made of platinum against platinum-rhodium alloy. This mapping permits an estimate of the small e.m.fs. due to the Benedick effect previously ignored. Since our values of e.m.f. are precise to 0·5 μV. we find it quite obvious that the Benedick effect does produce detectable values of e.m.f. If we want to get the pure modification of the Thomson gradient alone, this Benedick effect should be calculated and accounted for, at least approximately.

This is particularly marked in the phenomena near the Curie point. According to Broili's results, any increase in the maximum temperature after it has already passed the Curie point (near 350° C.) is accompanied by no further increase in thermomagnetic e.m.f. This indicated that the Thomson potential-gradient is no longer modified by the magnetic field in those parts of the specimen which lie above the Curie-point temperature. In the present work the same is true, but only approximately. Under zero tension, for instance, as the maximum temperature is increased from 320° C. to 380° C. there is already an increase of 1·0 μV., while a further increase of temperature to 500° C. produced a further e.m.f. of about 5·0 μV. The results when the tension was 8 kg. were much more marked; the change from 320 to 380 being accompanied by an increase in e.m.f. of 4·5 μV. We might explain these results as due to a shift of Curie point with tension, but the known dependence is in the opposite direction$^{(5)}$ and much smaller than that required for the present problem. We prefer to explain the e.m.f. as due to the temperature-gradient changing as the maximum temperature is raised. The Benedick e.m.f. on the two sides of the temperature map—one e.m.f. occurring in wire at zero tension and zero magnetic field, and the other in magnetized strained wire—will both be changed by increase in the temperature-gradient, and not in general by the same amount. Judged from the results of Pi and Band already mentioned, the effect is of the right order.

---

* Pure nickel wire manufactured by Schering-Kahlbaum AG.

§ 2. PROCEDURE AND APPARATUS

Figure 1 shows the arrangement. The test wire was stretched between two spring dynamometers, the clamps being simple screws with rubber washers bearing on the test wire. Microscope micrometers at either end, focused on two marks on the wire, permitted measurement of the elongation of the specimen and a check on its stability of position. The set-up was mounted on a heavy wooden stand which supported also the two water-cooling tubes and the heating-system symmetrically on either side of the central clamp. This clamp was rigidly made of brass casting with milled beds for the jaws, which were of snug fit; this eliminated all detectable shift in their position when they were subject to horizontal forces. The jaws were 0·5 cm. in width along the direction of the wire, from which they were electrically insulated by mica sheets. The slate heaters abutted immediately upon both sides of the clamp jaws. They were non-inductively wound with eureka wire to the same total resistance and were connected in series. An actual test of the eureka wire gave its composition as copper 55 per cent, nickel 40 per cent and zinc 5 per cent. This was the least magnetic resistance-wire available; a check on its permeability indicated even less magnetism than would be expected from the percentage of nickel present in it, but it is admitted that this may be a source of error, particularly in regard to the form of the hysteresis loops obtained, and in partially disturbing the uniformity of the magnetic field about the specimen.

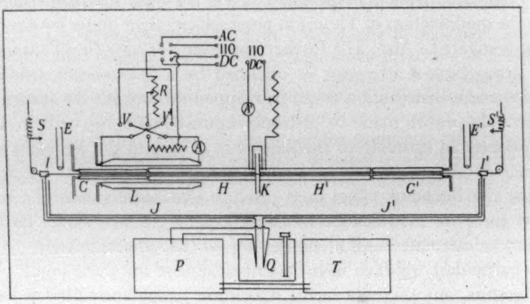


Figure 1. Arrangement of apparatus.

A, *A′*, ammeters. *C*, *C′*, water-cooling tubes. *E*, *E′*, microscope micrometers. *H*, *H′*, electrical heaters. *I*, *I′*, insulated clamps. *J*, *J′*, shielding tube round specimen. *T*, thermostat for standard cell. *V*, *V′*, special switching device for hysteresis loop. *K*, central clamp. *L*, water-cooled solenoid. *P*, K-type potentiometer. *Q*, oil bath with stirrer. *R*, resistance control to stopping point in hysteresis loop. *S*, *S′*, spring dynamometers.

The temperature-distributions produced by different working currents were approximately mapped by thermocouple probes used in preference to couples annealed to the test wire for reasons thoroughly discussed elsewhere$^{(6)}$. The element platinum and an alloy, platinum (90 per cent) + rhodium (10 per cent), gave a linear calibration curve to above the maximum temperature to be observed. One such couple was kept at a fixed position in the centre of each heater to check the constancy of the temperature-distribution during both the mapping and the subsequent thermomagnetic work. During the thermomagnetic observations continual control of the heating-current to a constant value was employed to ensure constancy of the temperature-distribution, the probes being removed from the apparatus after the mapping. Four temperature-distributions actually used are shown in figure 2.

The magnetizing solenoid of the Moullevigen form was so placed that the uniform field was superposed on that part of the wire which was subjected to the greatest temperature-gradient; it extended from the minimum temperature in the water-cooled part to the point of maximum temperature.

The magnetizing cycle was controlled by a switching-device made after the design of A. W. Smith(7), which carried the current through a complete cycle and stopped it at any desired point ranging from the maximum (10 A.) to as small a value as we pleased; actually 0·1 A. was the minimum used for the hysteresis work. This device is of fundamental usefulness, permitting as it does complete elimination of errors due to zero-wander and departure from the cyclic state of the specimen during measurement of the hysteresis loop. The loops actually obtained were perfectly symmetrical, never departing from symmetry by more than the 0·5 μV., which was the minimum difference detectable with certainty on the potentiometer used. By a few rapid oscillations of the switch between individual readings the specimen was maintained in the cyclic state as long as was required.

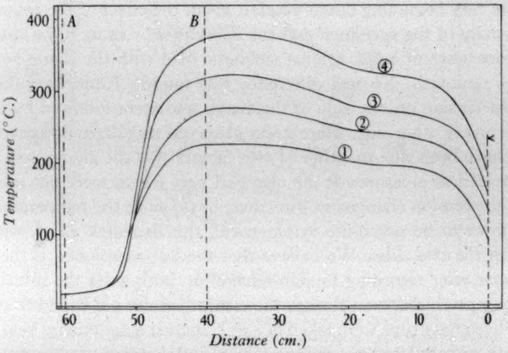


Figure 2. Temperature-distributions. The numbers correspond to those referred to in later diagrams of thermomagnetic hysteresis, etc. *A–B* is the extent of the uniform field.

The specimen wire extended from the tension-clamps at either end to an oil bath kept at uniform temperature by gentle stirring. Copper-wire connections to the K-type potentiometer completed the essentials of the electrical circuit.

In spite of all care in the construction of the apparatus, the temperature-distributions actually obtained were not perfectly symmetrical, although they were nearly so. Usually it was found that even when no tension and no magnetic field are applied, a small but definite e.m.f. was measurable on the potentiometer, presumably on account of the lack of symmetry. This e.m.f. was usually of the order of 2 or 3 μV., and was observed to vary in magnitude so long as the heating-currents were not kept exactly constant. It was found possible, however, by constant watch over the heating-system, to maintain this initial e.m.f. constant after the steady state had been reached. Since it had proved quite sensitive to slight changes in the temperature-distribution, it was therefore assumed that the temperature had remained constant throughout a set of observations if it was reproduced after the end of the set. If the initial e.m.f. varied by more than 0·5 μV. during a set of observations in spite of constant watch over the heating-system, then the set was rejected. Only seldom did this occur after the equipment had been brought into proper working order.

From the very beginning doubt was felt about the effect of the central clamp on the homogeneity of the specimen and the subsequent e.m.fs. to be studied. Runs were therefore taken of e.m.f. against magnetic field with the clamp both in action and entirely removed. No real difference was found. Runs were then taken of e.m.f. against tension on one side of the clamp and were followed by similar runs with tension on the other side. Here it was observed that slight differences occurred which may have been due to either of two factors: (*a*) the clamp may have introduced asymmetrical pressures at the clamped part which were not reliably duplicated when the tension changed in direction; or (*b*) since the temperature-distribution was known to be not quite symmetrical, the Benedick e.m.f. would change differently on the two sides. We believe that the last-mentioned is the real cause, chiefly because after returning to zero tension on both sides the initial e.m.f. was satisfactorily reproduced even though the clamp had not yet been released.

Finally, hysteresis runs were taken of e.m.f. against magnetizing field at different tensions, both with and without the use of the central clamp. In almost all cases the results were less consistent when the clamp was not in use; but otherwise there was no real difference between the two sets of results. The lack of mutual consistency among the results when the clamp was released must be due to slight shifts of the specimen permitted by the somewhat non-rigid clamps at the dynamometers. It was therefore decided that the use of the central clamp was justified for the purposes of (i) investigating the e.m.f. produced by a difference of tension in the two parts of the specimen and (ii) keeping the wire fixed mechanically during other observations in which tension was required in the part of the wire suffering magnetization.

The chief drawback to the use of the central clamp was that it necessarily caused a considerable dip, which could not be measured very precisely, in the temperature curves. This point must be considered when the distribution is analysed for the Benedick correction terms.

§ 3. RESULTS

Thermomechanical e.m.f. and hysteresis. The results shown in figure 3 were obtained by keeping the magnetic field zero and the tension on one side of the central clamp zero. The e.m.f. is plotted as a function of the tension on the other side; it is to be regarded essentially as due to a modification of the homogeneous thermoelectric constants in that part of the wire which is subjected to the longitudinal strain, which thus destroys the balance of the complete circuit. The total e.m.f. obtained was of the order of 20 μV., most of which must be due to modification of the Thomson potential-gradient. However, from results not yet published we estimate that several microvolts will be due to modification of the other constants a and b mentioned earlier. In particular there is no true saturation of the e.m.f. as the maximum temperature passes the Curie point, and the increase is not negligible in comparison with the total e.m.f.

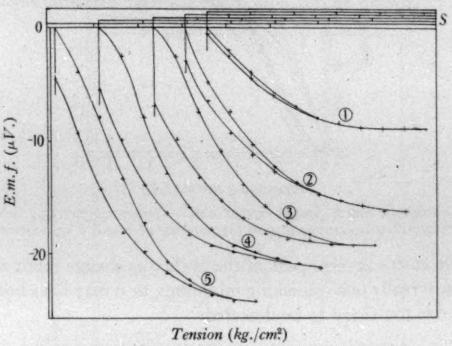


Figure 3. Thermomechanical e.m.f. and hysteresis. The intervals on the horizontal S axis represent* tension-increases of 2 kg. in every case; the hysteresis loops are drawn with their origins displaced for the sake of clarity. Curve (5), as in succeeding diagrams, corresponds with a temperature-distribution (not shown in figure 2) whose maximum was about 500° C.

The hysteresis observed is normal in form, and seems to increase with increasing temperature—a circumstance which is probably of some significance for the explanation of the hysteresis. If the latter were simply elastic lag, it would be expected to increase with rising temperature.

Thermomagnetic e.m.f. and hysteresis. The initial rise of e.m.f. against increasing magnetizing field is shown for the different temperature-distributions studied, in figure 4, for zero tension (and for a tension of 8 kg. for comparison later). Their lack of saturation as the temperature passes the Curie point has already been mentioned; otherwise they are saturation curves very similar in form to the known curve for the square magnetization plotted against field for nickel.

The hysteresis found for zero tension is shown in figure 5. There is very little difference between the various curves at different temperature-distributions. They are also similar to the curves obtained by Broili referred to previously. It is to be noted however that at no point in the cycle could the e.m.f. actually be brought back to the initial zero value, until of course after the final demagnetization. For the highest three temperatures the curves might be extrapolated to intersect near the zero axis, but the actual experiment failed to reveal such a point; the minimum e.m.f. obtainable at this inverse peak of the cycle was always about 2 μV. This may have been due to a really non-vanishing minimum, or it may have been due to a zero minimum that was too sharp to be detected.

---

* A tension of 1 kilogram corresponds to 127·3 kg./cm^2

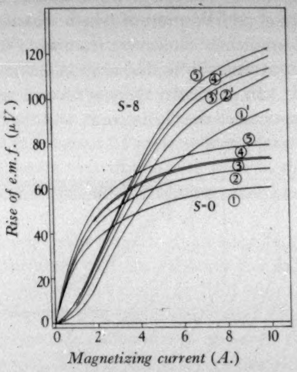


Figure 4. Thermomagnetic e.m.f. Initial rise of e.m.f. against magnetizing field H, for the five temperature-distributions indicated for tensions of 0 and 8 kg. respectively.

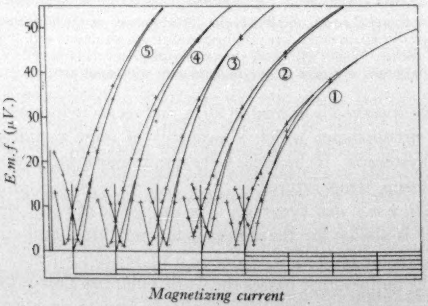


Figure 5. Thermomagnetic hysteresis loops. The intervals on the horizontal H axis represent 2 A. in every case*. The loops have been drawn with their origins displaced horizontally for clarity.

Mechanical-magnetic equivalence. From the above it will be seen that the e.m.f. produced by longitudinal strain is opposite in direction to that due to longitudinal magnetization. This is in harmony with the known fact that longitudinal strain produces radial magnetization in nickel. It was thought worth while to take a few runs of tension against magnetic field so balanced that the e.m.f. produced should remain at the initial zero. The results are given in figure 6. It is possible to assume that the two disturbances, because they annul one another, are essentially equivalent in magnitude, and this may furnish us with a good method of finding quantitatively the magnetizing effect of longitudinal strain. It is pointed out, however, that there is probably a skin effect in both the longitudinal and the transverse magnetizations: the former is already well-known, and the latter must exist to prevent infinite concentration of polarity down the axis of the wire. If the two skin-depths are almost equal, no magnetization will be produced when the tension and magnetic field are balanced in the above manner, and they can be truly considered as equivalent. If the skin-depths are different no such conclusion can be drawn.

---

* A magnetizing field of 10 amperes corresponds to 303 gauss.

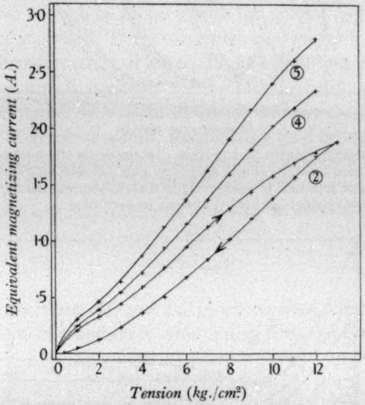


Figure 6. Mechanical-magnetic equivalence. Curves of magnetizing field H against tension S balanced for zero thermo-e.m.f. A complete hysteresis curve is shown in the case of lowest temperature-distribution; the numbers refer to temperatures previously mentioned.

The hysteresis curve shown in the same figure is of a form which evidently indicates that the radial magnetization follows the tension-change with less hysteresis than that with which the longitudinal magnetization follows the magnetizing field*.

Effect of strain on thermomagnetic hysteresis. Not only is the magnitude of the e.m.f. produced by magnetization changed when the wire is subjected to tension, but also the form of the hysteresis loop is greatly modified. This phenomenon is shown in figure 7. The variation of initial e.m.f. under zero field appearing in this diagram is of no significance; it arises merely from the fact that only half of the wire is stretched, the half not in the magnetizing solenoid being free from tension. The dotted curve passing through the initial points is simply the pure thermomechanical e.m.f. for rising tension already shown in figure 3. Had the other half of the wire also been stretched to the same tensions in each case this e.m.f. would have been eliminated, but obviously no change would have thereby been made in the form of the hysteresis loops. As regards this hysteresis itself it will be noticed that the area

---

* See additional note on demagnetization phenomena.

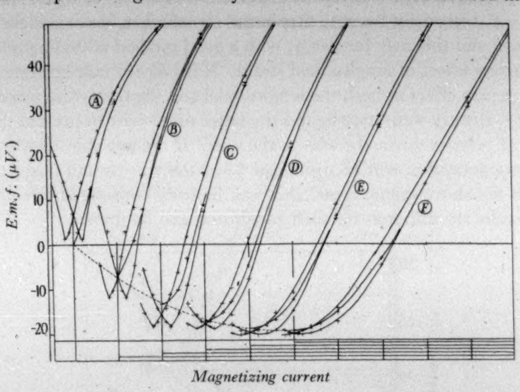


Figure 7. Thermomagnetic hysteresis and tension. Intervals on the horizontal H axis are in steps of 2 A. A refers to zero tension; B to 2 kg.; C to 4 kg.; D to 6 kg.; E to 8 kg.; and F to 10 kg. Only the lower part of one-half of each cycle is shown here to permit the use of a sufficiently large scale. Temperature-distribution (3) obtains in each case.

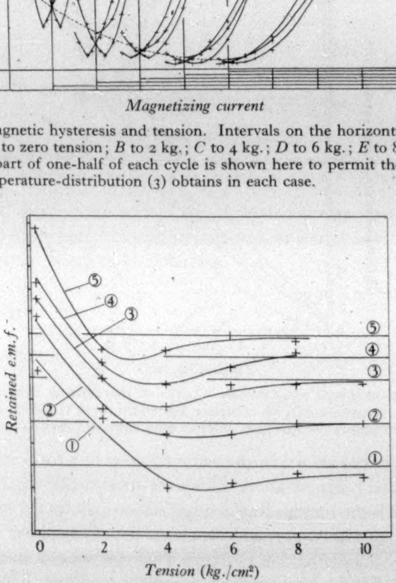


Figure 8. Retentivity of thermomagnetic e.m.f. The values of the retained e.m.f. relative to the initial e.m.f. at the beginning of the magnetizing cycle are shown for the four temperature-distributions indicated. The vertical axis shows intervals of 2μV. and the zero axis for each curve is shown by the correspondingly numbered horizontal line.

of the loop seems to increase a little for greater tensions, but the most obvious change is in the general shape of the curves. For quite small tensions the point of minimum e.m.f. in the hysteresis becomes greatly negative, but the minimum flattens out for greater tensions. In figure 8 are shown curves of retentivity against tension for the different temperature-distributions; in every case the retentivity changes from the normal positive anomalous negative values for small tensions returning asymptotically towards zero for high tensions.

§4. INTERPRETATION OF RESULTS

To a first approximation only, the electromotive forces observed in the present work can be considered as due to modifications of the Thomson potential-gradient in that part of the wire which is subject to the disturbances considered, namely longitudinal magnetization and longitudinal strain, or a combination of both. This approximation, as has already been emphasized, ignores the homogeneous e.m.f. associated with the square and cube of the temperature-gradient. The present results are quite precise enough to proceed with the second approximation in which these new e.m.fs. are included. This analysis is reserved for a later paper, pending completion of an exhaustive study of the constants a and b* on the lines begun in the work of Pi and Band already quoted.

a, b

To a first approximation, therefore, all the curves here obtained are really curves of Thomson potential-gradient against magnetizing field or tension as the case may be, except that the initial undisturbed value remains unknown or the position of the true origin of the curves is not determined.

Now the Thomson e.m.f. E produced in a wire between two points A and B will be given by

E

$$E = \int_A^B \sigma dT.$$

σ

If the wire is in two sections AB and BC in which σ has different values σ_1 and σ_2, and if the temperature-distribution is symmetrical with respect to the point B, then

$$E = \int_A^B \sigma_1 dT + \int_B^C \sigma_2 dT$$

$$= \int_A^B (\sigma_1 - \sigma_2)\, dT.$$

In general σ will be a function of both H, the magnetizing field, and S the tension. Also, if we write $\sigma\,(HS)$ to indicate this functional relation, we shall have

H, S

$$-\sigma\,(H_1S_1) + \sigma\,(H_2S_2) = \Delta\sigma,$$

also a function of temperature. In other words, the modification of the Thomson potential-gradient produced by a given ΔH or by a given ΔS will depend considerably upon the temperature T. Thus in setting the e.m.f. E, since

T

$$E = \int_A^B \{\sigma\,(H_2S_2) - \sigma\,(H_1S_1)\}\, dT$$

$$= \int_A^B \Delta\sigma . dT,$$

where S_1 and H_1 are usually both zero in our particular work, we cannot assume that E will be directly proportional to the temperature-difference $T_B - T_A$. However, since $\Delta\sigma$ is supposed to be a function only of temperature, and not of position in the wire (if the magnetization and tension are uniform), we can consider E to be a function only of $(T_B - T_A)$, and independent of the temperature-gradient in the wire.

---

* See page 905.

In particular if the temperature T_B is increased to T_B', say, a shorter part of the wire now exhibits the previous total e.m.f. E, while the part of the wire between the temperatures T_B' and T_B will now contribute a further e.m.f. given by

$$E' - E = \int_B^{B'} \Delta\sigma dT.$$

Turning to the curves shown in figure 4, we see that the temperature-difference $T_B' - T_B$ as between adjacent curves is not too great to prevent an average tempera-ture from being adopted for the range T_B to T_B', and then an average value $\overline{\Delta\sigma}$ of $\Delta\sigma$ for that range. We may then consider the change of e.m.f. as we pass from one temperature-distribution to the next, keeping H and S the same, to be given by

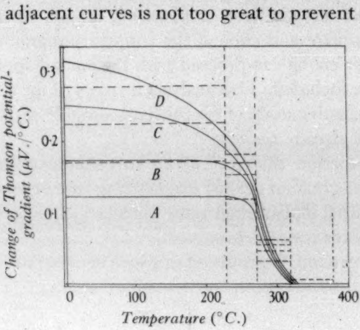


Figure 9. Thomson-potential susceptibility. The average change of Thomson potential-gradient due to fields (A) of 1 A., (B) of 2 A., (C) of 5 A., and (D) of 10 A., with the wire under zero tension*. Vertical dotted lines indicate temperature-ranges over which averages were taken, the latter averages being shown by horizontal dotted lines.

$\overline{\Delta\sigma}$

$$E' - E = \int_B^{B'} \Delta\sigma dT = \overline{\Delta\sigma}\,(T_B' - T_B).$$

By choosing successive steps we can thus obtain a curve of $\Delta\sigma$ against average temperature $\frac{1}{2}(T_B' + T_B)$ for each value of H and S. The initial step for the whole temperature-range of the lowest temperature is too big to give reliable averages, but it can be treated in the way adopted above for the purpose of providing a rough guide as to the behaviour of $\Delta\sigma$ at lower temperatures. The curves obtained in this way are shown in figure 8 for zero S, and in figure 9 when $S = 8$ kg.

It will be observed that $\Delta\sigma$ for any given H becomes zero near the known Curie point, at least for zero tension. There is an apparent shift of Curie point to higher temperatures when the tension is increased. This, as has already been mentioned, is interpreted as due to the Benedick effect ignored in this first approximation.

---

* A magnetizing field of 10 amperes corresponds to 303 gauss.

Further analysis in an attempt to get complete and precise dependence of $\Delta\sigma$ on T and of σ on H and S is therefore reserved until data on a and b are available.

The change in Thomson potential-gradient $\Delta\sigma$ (and thus the e.m.f. observed) is taken as positive when the positive current would, were the outside circuit closed, flow up the temperature-gradient in that part of the wire where the magnetic field or tension is causing the change in σ.

§ 5. NOTES ON DEMAGNETIZATION PHENOMENA

In the work on thermomagnetic hysteresis as affected by strain it was noticed that demagnetization destroyed not only the magnetism due to field H, but also that due to tension S. Thus, suppose E_0 to be the initial e.m.f. under zero field, so that $H=0$, and zero tension so that $S=0$. The hysteresis loop against H is obtained for some definite tension, say S_1, and afterwards demagnetization is carried out while $S=S_1$; it is found that E becomes E_1 which is $<E_0$. Releasing the tension, $S \to 0$, increases E_2 towards E_0, and a further demagnetization brings E back again to E_0 precisely. For a single demagnetization to give a direct check of E_0 it is necessary to release the tension beforehand.

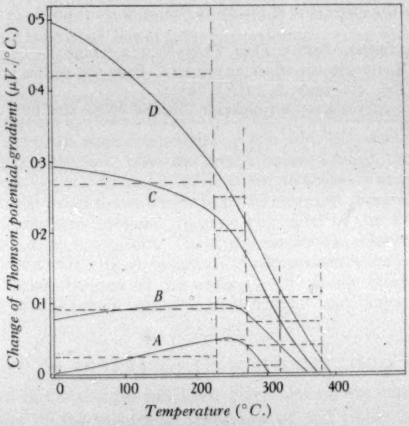


Figure 10. Thomson-potential susceptibility. The constitutions are the same as for figure 9, but the wire is under a tension of 8 kg.*

The magnitudes of E_0-E_1 and E_0-E_2 are of the same order as the values of E shown in figure 3 for the pure thermomechanical e.m.f.

From the above facts we may deduce the following interesting conclusions. (*a*) The radial magnetism produced by S_1 is destroyed by the usual demagnetization even while S_1 is mechanically maintained. (*b*) Even though this magnetism has been destroyed, stretched wire is still thermoelectrically different from unstretched wire;

---

* A tension of 1 kilogram corresponds to 127·3 kg./cm^2

otherwise E_1 and E_0 should be identical. (*c*) On releasing the tension in the demagnetized wire radial magnetism reappears with sense opposite to that caused ordinarily by that tension: because the change E_1 to E_2 is in the positive direction, whereas the e.m.f. produced by tension is negative.

It was also noticed that the e.m.f. retained after a pure tension hysteresis cycle, figure 3, was removed by the usual demagnetizing process. This might be explained as due to magnetostriction vibrations during demagnetization if the hysteresis is regarded as purely mechanical in origin.

REFERENCES

(1) Heinz Broili. *Ann. d. Physik*, **14**, 3 (1932).

(2) Chang and Band. *Proc. Phys. Soc.* **45**, 4 (1933).

(3) Pi and Band. "Longitudinal thermoelectric effect: (2) nickel in longitudinal magnetizing field." Page 852 of this volume.

(4) Thomson. See the statement of theory in the work on copper. *Proc. Phys. Soc.* **46**, 3 (1934).

(5) D. P. Ray-Chaudhuri. *Zeits. f. Phys.* **71** (7–8), 473–7 (1931).

(6) "Longitudinal thermoelectric effect: (5) silver." Page 904 of this volume.

(7) A. W. Smith. *Rev. Sci. Instr.* **5**, 3 (Oct. 1932).

(8) See for instance curves given for nickel in *Applied Magnetism* by T. F. Wall (E. Benn, Ltd.), p. 88.

55.037.2 631.437.1

THE ELECTRICAL PROPERTIES OF SOIL AT FREQUENCIES UP TO 100 MEGACYCLES PER SECOND; WITH A NOTE ON THE RESISTIVITY OF GROUND IN THE UNITED KINGDOM

By R. L. SMITH-ROSE, D.Sc., Ph.D., A.M.I.E.E.,
National Physical Laboratory

Received April 2, 1935. *Read June* 21, 1935

ABSTRACT. Previous papers have described the use of a laboratory method for investigating the electrical properties of soil under alternating-current conditions at frequencies ranging from 50 c./sec. up to 10 Mc./sec. The measurements have now been extended to 100 Mc./sec., and the results are described in the present paper. Substantially the same method has been used, but in order to provide an overlap with the previous work, the present measurements cover the frequency range of 1 to 100 Mc./sec. Check measurements carried out on small fixed resistors indicate that the overall accuracy at the highest frequencies is better than 20 per cent. As a result of this work it may now be stated that for normal samples of surface soil taken from the National Physical Laboratory, the conductivity is of the order of 10^8 e.s.u. at all frequencies up to 1 Mc./sec., rising to rather less than twice this value at 100 Mc./sec. Over the same frequency range, the dielectric constant decreases from about 10^5 at a frequency of 50 c./sec. to about 15 at 100 Mc./sec.

An appendix to the paper draws attention to the electrical-resistivity maps of England and Wales and of Southern Scotland, recently published by the British Electrical and Allied Industries Research Association. These maps indicate in coloured form the apparent resistivity of the ground for a depth of 500 ft. A comparison is made between the values so indicated and those obtained by the author in previously described measurements on samples of soil obtained from various sites in England and Wales.

§ 1. SCOPE OF INVESTIGATION AND METHOD ADOPTED

CONSIDERABLE attention has been given during the past few years to investigations of the electrical properties of soil under alternating-current conditions. Previous papers(1, 2) have described the use of a laboratory method for measuring these properties and the results obtained on a variety of types of soil over a frequency range from 50 c./sec. to 10 Mc./sec. (wave-length 30 metres). The present paper describes the extension of these measurements to a frequency of 100 Mc./sec. (wave-length 3 metres).

For purposes of clarity it will be an advantage to recapitulate briefly the principles of the method employed. A condenser of suitable size, shape and capacitance is filled with a sample of the soil under examination, and its capacitance and effective shunt resistance are measured at the required frequency. From the values so obtained and a knowledge of the capacitance of the condenser with air as the dielectric, the specific conductivity and dielectric constant of the soil can be ascertained.

§ 2. APPARATUS AND EXPERIMENTAL PROCEDURE

The measurements at frequencies up to 10 Mc./sec. were made by using a concentric cylindrical condenser, a section of which is indicated in figure 1 (*a*). The principal dimensions of this condenser are indicated in the figure, and its capacitance with air as the dielectric was found to be 2·5$\mu\mu$F. For the higher frequencies the physical dimensions and capacitance of this condenser make its use difficult, and the measurements at frequencies between 10 and 100 Mc./sec. were accordingly made in the small parallel-disc condenser depicted in figure 1 (*b*). The capacitance of this condenser, with air as the dielectric, was calculated to be 0·15 $\mu\mu$F.; and this value was checked by measuring the capacitance when the condenser was filled with distilled water.

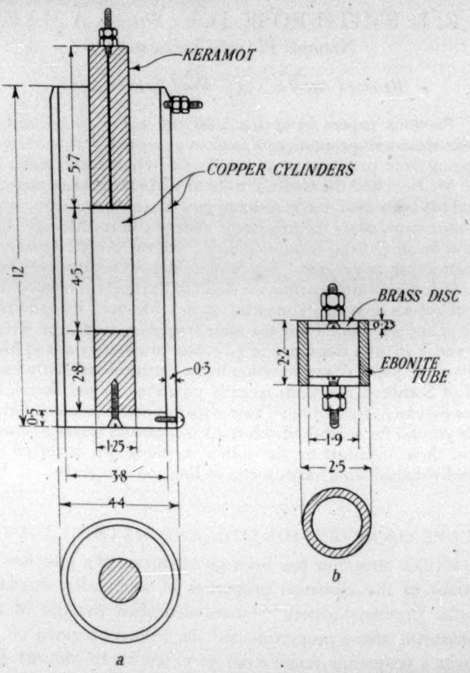


Figure 1. Diagrams of condensers used for soil measurements. (*a*) Large container used for frequencies up to 10 Mc./sec. (*b*) Small container used for frequencies of 10–100 Mc./sec.

For measurements on soil the appropriate condenser was filled with the sample under examination, and was connected in parallel with a resonant circuit comprising a small inductance coil of a few turns and a calibrated variable condenser of maximum capacitance about 20 $\mu\mu$F. At the highest frequency of 100 Mc./sec., the dimensions of the resonant circuit were reduced as much as possible: a single-turn loop 8 cm. in diameter was connected directly across the variable condenser, with all leads reduced to the minimum. The supply of oscillatory current for the measurements was obtained from a valve oscillator of reasonably good frequency-stability. An oscillating detector-amplifier was used to provide, by the heterodyne beat method, a signal audible in a pair of telephones. It was found to be neither readily practicable nor necessary to employ the double-heterodyne-beat method which was used in the measurements at lower frequencies$^{(2)}$.

In carrying out a series of measurements, the tuned circuit alone was first coupled to the oscillator and adjusted to resonance. The variable tuning condenser was then altered on either side of the resonance position to ascertain the capacitance-change required to produce the maximum frequency-change of the oscillator. The effective shunt resistance R_0 of the circuit may then be calculated from the formula R_0

$$R_0 = \frac{1}{\pi f . \Delta C_0} \qquad \text{......(1)},$$

where ΔC_0 is the total change in capacitance for the maximum frequency-change ΔC_0
and f is the frequency. f

The sample condenser full of soil was then connected across the circuit, and the above operations were repeated to find the new value of capacitance-change, ΔC_1. ΔC_1
The resistance of the condenser full of soil may then be obtained from the relation

$$R_s = 1/\pi f (\Delta C_1 - \Delta C_0). \qquad R_s$$

The change in the adjustment of the variable condenser in the two resonance conditions was due partly to the value of the added capacitance and partly, and to a smaller extent, to the effect of the resistance added to the tuned circuit. Allowance could be made for this latter effect so that the capacitance of the condenser full of soil was known, and from this the effective dielectric constant of the sample of soil under examination was obtained.

§ 3. PRECAUTIONS OBSERVED AND CHECKS ON METHOD

It will have been noted from the previous section that the measurements are made in terms of capacitance-changes of a variable air condenser. It is therefore necessary to know with some precision the effective capacitance of the condenser between its terminals. If C is the static or low-frequency capacitance of the con- C
denser, and L is the residual inductance measured from its outside terminals, then L
the effective capacitance C_e at any frequency $\omega/2\pi$ is given by the relation C_e

$$C_e = \frac{C}{1 - \omega^2 LC} \qquad \text{......(2)}.$$

For most ordinary condensers used at low radio frequencies the quantity $\omega^2 LC$ is reasonably small, but at high frequencies even a good-class condenser with a reasonably low residual inductance may have an effective capacitance which differs considerably from its low-frequency value, particularly when the value of C itself is large. In the case of the larger of the two variable condensers used in these measurements, the value of L was found to be $0{\cdot}04\,\mu$H., so that at maximum scale reading the ratio of C_e/C was 1·13 for a frequency of 60 Mc./sec. This error in the apparent or low-frequency capacitance was allowed for throughout the measurements. The other condenser had such short leads to the inductance coil that their effect was considered to be negligible even at the highest frequency employed.

For the purpose of accurately locating the soil condenser, small mercury cups were provided in the stout leads connecting the variable condenser to its inductance coil. These cups were arranged to receive the very short leads from the terminals of the soil condenser, or alternatively from small fixed resistances used for checking purposes. Care was taken throughout the measurements to avoid the influence of stray capacitances and the effect of any body movements of the observer.

In order to obtain some check of the accuracy of the resistance values obtained by this method at very high frequencies, measurements were made on a number of small fixed resistors of the right order of magnitude. These resistors were of commercial pattern and of four different types, varying from a thin film deposited on a tube 8 mm. in diameter to solid rods of material of from 1 to 10 mm. in diameter. Table 1 shows the results of a series of measurements on these resistors by the method described above. In a few typical cases the values of the resistance measured at 1 Mc./sec. are included.

Table 1. Measurement on four types of resistors at various radio frequencies

Resistor type	Measured values of resistance at frequencies (Mc./sec.)				
	d.c.	1	30	60	100
1	1020	—	930	990	860
	2060	—	1730	1930	1900
2	1000	995	940	970	840
	2070	—	1860	2080	1900
3	1030	1030	950	970	820
	2150	—	1790	2250	2100
4	1030	1030	970	990	910

The largest error in these measurements is 20 per cent, and in the majority of cases the discrepancy between the d.c. and a.c. values is less than 10 per cent. This discrepancy includes both the error of the measurement and any change in value of the resistor at high radio frequencies from its direct current value. Since, as has already been remarked, the resistors were of very different types, they would not be expected to show a consistent change in their values at the higher frequencies, which might influence the above results in the same direction for each type. As it was considered that an accuracy of 20 per cent was adequate for the measurements on soil, the above discrepancies were not investigated in any great detail.

It is to be observed in the above table that there is a tendency for the radio frequency value of the resistance to be less than the direct current value. In the case of a resistor of type 4 having a direct-current resistance of 1990 Ω., the value obtained at the high radio frequencies was about half this figure. This discrepancy was traced to the fact that a resistance film was deposited on both the inner and outer walls of a porcelain tube, but that conductive connexion from the terminals was made only to the outer film. At very high frequencies, however, the effect of the capacitance between the films was to connect them in parallel with each other, so as to give an effective resistance approaching half the value obtained in a direct-current or low-frequency measurement. On removal of the inner film, the value of the resistance was found to be 1930 Ω. at the frequency of 60 Mc./sec. It is possible that some of the other errors recorded in the above table are due to parallel-resistance effects of an analogous type.

§ 4. MEASUREMENTS ON SOIL

By means of the experimental procedure outlined above, measurements were made on samples of soil taken from a site at the National Physical Laboratory, which had been used as site 1 in previous work⁽¹⁾. Some check measurements were carried out at a frequency of 10^6 c./sec. in order to establish continuity with the results recorded in previous papers. In these cases the measurements were made both in the small condenser used for the present frequency range and also in the larger condenser used in all the previous work. Some care was necessary in packing the smaller condenser to the required degree, but the agreement obtained between measurements at the same frequency in the two condensers indicates that this packing was satisfactorily accomplished. As in the previous work, the moisture-content of the soil was determined from the loss of weight on drying it out at a temperature of 100° C. The values of the moisture-content included in table 2 are referred to the weight of the dry soil.

Table 2. Summary of measurements on soil from site no. 1 at the National Physical Laboratory

Sample No.	Moisture-content (per cent)	Frequency (Mc./sec.)	Conductivity*		Dielectric constant	
			In small container (e.s.u. × 10^{-8})	In large container (e.s.u. × 10^{-8})	In small container	In large container
A	24	1·0	0·9	1·0	55	55
		10	1·3	—	25	—
		60	1·3	—	18	—
		100	1·5	—	15	—
B	25	1·0	0·9	0·9	40	50
		10	1·1	—	21	—
		30	1·3	—	21	—
		60	1·5	—	20	—
		100	1·9	—	15	—
C	30	1·0	0·8	0·8	45	50
		10	1·0	—	25	—
		30	1·2	—	20	—
		100	1·3	—	20	—

* Conductivity (e.s.u.) = $9 \times 10^{11} \div$ resistivity (ohm-cm.).

These results show that for the samples of soil selected the conductivity increases progressively from about $0 \cdot 9 \times 10^8$ e.s.u. at a frequency of 1 Mc./sec. to a value of from $1 \cdot 3$ to $1 \cdot 9 \times 10^8$ e.s.u. at 100 Mc./sec. Over the same frequency range the dielectric constant falls from about 50 to a limiting value which appears to lie between 15 and 20. The agreement between the results obtained with the two condensers at a frequency of 1 Mc./sec. is considered to be quite satisfactory and to justify the linking up of the present results with those previously obtained at low frequencies. These showed that the value of the conductivity remained practically unchanged for all frequencies from 50 c./sec. up to 1 Mc./sec., although the dielectric constant decreases steadily from very high values at low audio frequencies.

§ 5. CONCLUSIONS

It may thus be concluded that for soil of normal moisture-content from the particular site selected at the National Physical Laboratory, the conductivity is of the order of 10^8 e.s.u. at all frequencies up to 1 Mc./sec., and then rises to rather less than twice this value at 100 Mc./sec. Over the same frequency range the dielectric constant decreases from about 10^5 at a frequency of 50 c./sec. to about 15 at 100 Mc./sec. The abnormally high values experienced at the lower frequencies have been discussed briefly in a previous paper⁽²⁾; they are probably attributable to the formation of a polarization film over the surface of the electrodes. In the case of soil, these high values of dielectric constant are of no practical importance at the frequencies at which they are obtained. It is only at frequencies above 1 Mc./sec. that the dielectric constant need be taken into account in most radio problems.

The variation of these electrical properties with the nature of the soil as obtained from different localities, with moisture-content and temperature has been previously investigated in some detail at frequencies up to 10 Mc./sec.⁽²⁾. It has not been considered useful at this stage to extend this study to the higher frequencies now covered.

§ 6. ACKNOWLEDGMENTS

This investigation was carried out as part of the programme of the Radio Research Board, and the paper is published by permission of the Department of Scientific and Industrial Research. The author is indebted to the able assistance rendered by Mr A. C. Gordon-Smith in connexion with measurements involved in the work.

REFERENCES

(1) R. L. Smith-Rose. "The electrical properties of soil for alternating currents at radio frequencies", *Proc. roy. Soc.* A, **140**, 359–77 (1933).

(2) R. L. Smith-Rose. "Electrical measurements on soil with alternating currents", *J. Instn. elect. Engrs.* **75**, 221–37 (1934).

APPENDIX

NOTE ON THE ELECTRICAL RESISTIVITY OF GROUND IN THE UNITED KINGDOM

The British Electrical and Allied Industries Research Association have recently published* electrical-resistivity maps of England and Wales and of Southern Scotland. These maps indicate in coloured form the apparent resistivity of the ground for a depth of 500 ft., and they have been prepared by Mr A. Broughton Edge for the above Association. During the past year or two the author has been investigating by a laboratory method the electrical properties of soil for alternating-current conditions at various frequencies from 50 c./sec. up to high radio frequencies. The results of such measurements made on samples of soil taken at various depths down to 10 ft. (in one case 300 ft.) from some 16 different sites in England and Wales were published in 1934†. It is interesting to compare these results with those indicated on the electrical-resistivity map referred to above.

The following table shows such a comparison for the 16 localities in question. The author has utilized the results obtained by him at a frequency of 1000 c./sec., and it is understood that the measurements by Mr Broughton Edge were made under direct-current conditions or with alternating-currents of various frequencies up to about 3000 c./sec. Measurements published in the paper referred to above have shown that, except in the case of the higher-resistivity samples of stone and

Electrical resistivity (kΩ.-cm.) of soil in different parts of England and Scotland

Site	Value according to the map by A. Broughton Edge for depth of 500 ft.	Values obtained by R. L. Smith-Rose for depth of 10 ft.
Rugby, Warwick	1 to 3	1·5
Baldock, Herts	3 to 30	6·9
Tatsfield, Kent	3 to 30	6·0
Brookmans Park, Herts	3 to 30	6·0
Daventry, Northants	1 to 3	35
Washford Cross, Somerset	1 to 3	2·1
Brendon Hills, Somerset	30 to 300	$3·5 \times 10^4$
Merrivale, Devon	>300	6×10^5
Dousland, Devon	>300	3×10^5
Moorside Edge, Yorks	3 to 30	29
Westerglen, Falkirk	3 to 30	4·3
Teddington, Middlesex	<1	7·5
Droitwich, Worcester	1 to 3	2·4
Slough, Bucks	1 to 3	4·0‡
Banstead, Surrey	3 to 30	4·5‡
Pulham, Norfolk	3 to 30	3·5‡

‡ Soil taken within 1 foot of surface and measured at 100 kc./sec.

---

* Shown at the Physical Society's Exhibition, January 1935.

† See reference (2) above.

slate, there is no substantial change in resistivity of soil in changing the frequency from 50 to 1000 c./sec.

The table shows that in 12 cases out of the 16 cited there is good agreement between the two sets of measurements. In these 12 cases, moreover, the author's results show that except for surface soil, the values of resistivity obtained at various depths from 1 to 10 ft., all lie within the limits specified by the map. The note on the map stating that "it is unsuitable for obtaining the resistivity corresponding to shallow depths" may thus need some qualification.

The cases in which there is a discrepancy between the two sets of results are discussed individually below.

Daventry. There is a 10 : 1 discrepancy in the values of resistivity indicated in the table. It may be remarked that the ordinary geological map* shows streaks of middle and upper lias through the Daventry and Northampton region, as distinct from the lower lias at Rugby. This is reflected in the author's measurements, which show that the resistivity at Daventry, where the subsoil is sandy, is over 20 times as great as that at Rugby, where the subsoil is clay. No similar streaks appear on the B.E.A.I.R.A. map, which shows a large belt of the midlands in the same resistivity range.

Brendon Hills. In this case the author's measurements on all samples taken from depths of 1 to 10 ft. showed a resistivity greatly in excess of 300 kΩ.-cm., the maximum value ascribed to this region by the B.E.A.I.R.A. map. The samples comprised a mixture of loam and slate, the slate-content and resistivity increasing with the depth. It may be noted that for this part of the country the resistivity map follows very closely the geological map to which reference was made above.

Teddington. Measurements of the resistivity of the ground within a few feet of the surface at the National Physical Laboratory, Teddington, have been carried out for several years past by a variety of methods and at various frequencies from 50 c./sec. up to 10^8 c./sec. In all cases the results have indicated a resistivity in excess of 2 kΩ.-cm., and generally above 5 kΩ.-cm. The probable explanation of the much lower value obtained by Broughton Edge is that it is due to the influence of the relatively low-resistivity clay bed which would be encountered at greater depths.

Slough. A similar effect may be influencing the results at Slough, although to a lesser extent on account of the proximity of this locality to the edge of the clay bed; for here again measurements of the ground resistivity for small depths at the Radio Research Station, Slough, have consistently given values in excess of 3 kΩ.-cm.

Conclusion. From this survey it appears quite definitely that the electrical resistivity map compiled by A. Broughton Edge and recently published by the B.E.A.I.R.A. is a valuable contribution to our knowledge of the resistivity of the ground in this country. It is possible that a reference to the detailed observations from which the map was compiled might elucidate one or two of the discrepancies pointed out above. Further, the available evidence shows that the map forms a safe guide to the resistivity values for alternating-currents of all frequencies from

---

* Geological Map of England and Wales by Sir Archibald Geikie.

50 c./sec. into audio range. Although the map is intended to indicate the resistivity corresponding to depths of about 500 ft., it would appear in many localities to be fairly representative even for depths of only a few feet. Further work on an extended scale would, however, be necessary before a definite recommendation could be made that the map should be used as a reliable guide to surface conditions.

DISCUSSION

Mr G. D. Pegler. The author's observation on resistors should be of value to workers at high frequencies. Since a comparison is made between the values of soil-resistivity measured by him for depths down to 10 ft., and the results indicated on the map for a depth of 500 ft., it would be interesting to know if he has made any investigation of the variation of soil-resistivity with pressure.

Dr L. Hartshorn. I am inclined to think that the low values of resistance obtained for the fixed resistors at very high frequencies are not altogether due to experimental error. I have observed such an effect even in resistors of the solid-rod type, and it is to be expected in all composite materials since they may be regarded as consisting of a complicated network of resistances and capacitances. The conductance of any such combination must increase with increasing frequency. The effect is well known in solid dielectrics. The soil and the composite resistor may be regarded as special cases, in which the effect becomes considerable in the range of frequencies used in this investigation.

Author's reply: In reply to Mr Pegler: I have in the course of my investigations on the electrical properties of soil made certain measurements on the effect of pressure on these properties. These measurements were, however, carried out merely with the object of ensuring that the laboratory tests should be representative of the actual conditions of the soil at the site under exploration. In the case of all results published, the sample condensers have been packed to a density approximately equal to that of the soil in its natural state.

Dr Hartshorn's comments are gratifying in suggesting that the experimental error is possibly somewhat less than that indicated in the paper.

548.73

THE QUANTITATIVE MEASUREMENT OF THE INTENSITY OF X-RAY REFLECTIONS FROM CRYSTALLINE POWDERS

By J. C. M. BRENTANO, D.Sc., F.Inst.P.

Received January 29, 1935. *Read in title May* 3, 1935.

ABSTRACT. The paper refers to determinations of the dispersion of F-values in the range of the L absorption levels, in which powder methods were applied to measurements requiring a high degree of accuracy and discusses the conditions determining quantitative measurements made with microcrystalline powders.

The absorption in a powder layer and the effect of the size of the crystallites on the evaluation of intensities are discussed. Indications are made as to the preparation of suitable particles and as to the use of spacing materials for avoiding the formation of larger units. A detailed discussion is given of the factors determining quantitative measurements by the method of the flat powder layer and by the method of mixed powders. The conditions determining the accuracy of the microphotometric photometry of reflections obtained from a flat layer and the absorption effect in a mixed layer in relation to the absorption in the individual particle are considered. Expressions are derived for evaluating the absorption effect for particles of different absorbing-power.

Reference is made to a recent paper by Brindley and Spiers on a development of the method of the flat powder layer and of the method of mixed layers. It is shown that this particular procedure which had been described by the writer in an earlier paper is not advisable for accurate measurements.

---

§ 1. INTRODUCTION

Recently a series of measurements have been made on the dispersion of X-ray scattering in the region of the L absorption levels by Baxter and the author(1), leading to the establishment of an F-dispersion curve for tungsten for the wave-length range from 0·49 A. to 2·3 A. The evaluation of the dispersion, which was derived from the differences of F-values obtained for different wave-lengths, required a high degree of accuracy; on the other hand particular difficulties were presented by the absorption in the tungsten, which is considerable for some of the wave-lengths used, and by the fact that the measurements had to be made for strong reflections in the range of small values of sin θ/λ where extinction effects are particularly large.

The measurements were made upon powders by determining photographically the intensities of the reflections scattered from a flat layer and using a mixture of powders(2), so as to obtain relative measurements in terms of a reference substance. These two methods, introduced by the writer in 1928 for quantitative evaluations from powders, have since been used in a number of instances. Intensity-measurements from powders present considerable difficulties; on the other hand they are the only way of approach to a number of determinations where large single crystals cannot be used, and we believe that a remarkably high degree of accuracy can be obtained if due precautions are taken. Discrepancies in particular with the method of powder mixtures in the hands of different observers seem to be due to the fact that proper regard had not been given to the conditions requisite to such measurements. A more comprehensive analysis should therefore be given, as the conditions determining the accuracy of powder measurements have only been discussed in an incidental way.

Brindley and Spiers have recently discussed in these *Proceedings*(3) a particular development of the methods. This had been described by the author in an earlier paper that probably escaped their attention; reference to this arrangement, which is not advisable for accurate measurements, will be made in the course of this paper.

§ 2. EXTINCTION EFFECTS IN POWDERS

One particular difference between the use of fine crystalline powders and single crystals for measuring the intensities of X-ray reflections depends on the fact that the size of the powder particles determines an upper limit to the size of the regularly spaced crystal units or crystallites. In this way by using a sufficiently fine powder the extinction effect of large single crystals, which Darwin describes as primary extinction and which in the first place impairs the interpretation of the intensities of reflections from large crystals, can be avoided. These conditions have been discussed by Darwin(4) and Ewald(5) and need not be recalled here.

Further, the coefficient of absorption enters into the expression for the scattered intensity. This has a different character for a fine powder and for an extended single crystal of mosaic structure. For the latter the absorption coefficient for a given wave-length in an angular range near the angle of reflection contains a term for secondary extinction, which increases as the setting for a strong reflection is approached in a manner depending on the angular spread of the blocks constituting the particular mosaic. For a powder of sufficiently small crystallites oriented at random an effect analogous to secondary extinction takes place in so far as the beam of wave-length λ going through the powder encounters crystallites so oriented λ
as to satisfy the reflecting condition

$$\sin \theta = \lambda / 2d_{(hkl)} \qquad \ldots\ldots(1),$$

where θ is the glancing angle of incidence and of reflection with respect to a set of θ
lattice planes (hkl) with the spacing $d_{(hkl)}$. All lattice spacings for which the spacing $d_{(hkl)}$
$d_{(hkl)} > \lambda/2$ contribute to this effect, which we have shown(2) to be a discontinuous function of the wave-length; with powder particles oriented at random it is however the same for all directions and independent of the particular reflection we observe. The powder offers thus the advantage that the same absorption coefficient enters into the intensity expression for all reflections and further that a greater number of reflections can be observed than with a single crystal, where the measurement is dependent on the presence of sufficiently well-developed crystal faces.

On the other hand the numerical value of the absorption coefficient varies greatly with the individual powder specimen; it depends on the closeness of the packing and on the extent to which the interspaces between the powder particles are filled with binding or spacing material; further the appearance of the full number of reflections leads to superpositions.

These points largely determine the procedure to be adopted in the interpretation of intensity-measurements from powders.

§ 3. SCATTERING FROM VOLUME ELEMENT OF POWDER

dV

In discussing the X-ray scattering from a micro-crystalline powder it is convenient to refer to the intensity of reflection from a volume element dV sufficiently small to produce no appreciable absorption and containing an adequate number of crystallites oriented at random so as to present no preferential orientation of lattice planes with respect to any direction.

The intensity of scattering is then given by

$$dP_{(hkl)}=\tfrac{1}{2}Ip_{(hkl)}\,F^2_{(hkl)}\,N^2q\cos\theta\, dV \qquad \text{......(2)},$$

$P_{(hkl)}$
I
$F_{(hkl)}$
N
θ
$p_{(hkl)}$
q

where $P_{(hkl)}$ is the total energy scattered in unit time for the reflection (hkl) out of a parallel incident beam for which I represents the intensity, i.e. the energy per unit time through unit cross section. $F_{(hkl)}$ is the scattering factor per unit cell (often the letter S is used for this term to distinguish it from the scattering factor of the atom) and N is the number of scattering cells per unit volume of the powder. θ is the glancing angle relating to the particular reflection and wave-length. $p_{(hkl)}$ is the multiplicity factor of the reflection, that is, the number of times the particular lattice appears with different orientation in the crystal form. The quantity q, where

$$q=\frac{e^4}{m^2\,c^4}\,\frac{\lambda^3}{\sin 2\theta}\cdot\frac{1+\cos^2 2\theta}{2}\,e^{-b\sin^2\theta/\lambda^2},$$

b

collects those quantities which establish the numerical link between I and $dP_{(hkl)}$ on the basis of classical theory, in accordance with the usual definition of the scattering-factor F. The exponential term $e^{-b\frac{\sin^2\theta}{\lambda^2}}$ relates to heat motion. No indices are given to θ and q, but they also refer to the particular reflection.

The scattering-factor $F_{(hkl)}$ which defines in the traditional way the scattering from the cell in terms of the scattering from a single free electron depends on the atomic scattering-properties of the atoms constituting the cell and on their spacial arrangement within the cell, leading to additive and subtractive interference effects. It takes account of the fact that only the coherent part of the scattered radiation contributes to the reflection and of the dispersion of X-ray scattering, by which the scattering from any atom depends on the relation between the wave-length and its energy levels. $F_{(hkl)}$ considered as an experimental quantity has no intrinsic value unless it is related to a certain wave-length.

In the case of crystals constituted of one kind of atoms only, as, for instance, metals, F can be introduced as the scattering-factor of the atom, when N becomes

Loschmidt's constant, the number of atoms in unit volume. Only those reflections of the simple lattice will appear for which the scattering contributions from the different atoms are additive according to the type of unit cell of the crystal. In our paper in 1928 the expression was applied to reflections from gold, silver and aluminium and was used in this sense.

§ 4. ABSORPTION EFFECTS IN FINE POWDERS

The scattering from an extended block of powder is obtained by integrating over its volume after multiplying the contributions from each volume-element by an absorption term

$$\exp\left(-\mu s_{xyz}\right) \qquad \ldots\ldots(3),$$

where $s_{(xyz)}$ is the total distance which the incident and scattered beam traverse in the powder as determined by the position of the point xyz within the powder. μ is the linear absorption coefficient of the powder; it acquires a definite value only when the specimen is constituted of such small particles distributed at random that it can be regarded as homogeneous for dimensions in which absorption becomes appreciable. We can then introduce a general coefficient of absorption μ which is related to the absorption coefficient of the constituents according to the volume they occupy.

$s_{(xyz)}$

μ

It can easily be shown that for two constituents 1 and 2 with the coefficients of absorption μ_1 and μ_2 we have

$$\mu = \frac{\mu_1 \, m_1 \, \rho_2 / m_2 \, \rho_1 + \mu_2}{m_1 \, \rho_2 / m_2 \, \rho_1 + 1} \qquad \ldots\ldots(4),$$

where m_1/m_2 is the ratio of masses and ρ_1/ρ_2 the ratio of densities.

The value of μ differs considerably from the absorption coefficient of the individual crystal measured within or outside the reflecting position and varies considerably with the constitution of the layer; for a given wave-length it will, however, have the same value for all reflections. The coefficient of absorption μ comprises any secondary extinction effect of the crystalline powder particles, to which we have referred.

When the intensities of scattering for various angles for one kind of crystals are compared, for instance when an F curve expressing the change of F with $\sin\theta/\lambda$ for a given wave-length is being determined, the uncertainty about the numerical value of μ for the powder can be overcome by choosing such conditions that for all angles of scattering the absorption term enters into the intensity expression as a constant multiplicative factor. This condition is not generally satisfied with the thin powder rod used in the Debye-Scherrer camera, where the scattering in different directions depends on the extent to which the deeper layers of the rod contribute to the scattered intensity. Definite conditions are obtained with the powder rod only when the absorption in the rod either is negligible or is so great that scattering is confined to a thin layer near to the surface. In the latter case for a given shape of rod the geometric location of the scattering volume becomes definite, so that the intensity-distribution can be calculated or established experimentally in a general way. This procedure, which is particularly adapted to very heavy atoms, has been introduced by Claassen⁽⁶⁾, and has been discussed in detail by Rusterholz⁽⁷⁾. When the depth of penetration into the layer has to be considered, approximate expressions apply, which depend, however, on the numerical value of the absorption coefficient.

§ 5. THE FLAT POWDER LAYER

The writer has discussed the reflection from a flat powder-layer⁽⁸,²⁾ and has shown that it can be used in a general way to obtain measurements of relative intensities independent of the absorption coefficient μ.

For a flat layer sufficiently thick to absorb fully the incident radiation, placed so that its normal is in one plane with the direction of incidence and with the direction in which the scattered beam is observed,

$$dP_{\zeta(hkl)} = \frac{I p_{(hkl)} F^2_{(hkl)} N^2 q f \, d\zeta}{8\pi\mu \, (1 + \sin\alpha/\sin\beta)} \qquad \text{......(5)},$$

$P_{\zeta(hkl)}$ where $P_{\zeta(hkl)}$ is the quantity of radiation scattered for the reflection (hkl) in unit
ζ time into the small azimuthal angle $d\zeta$ near the plane in which scattering is observed,
f, α, β when the parallel beam of intensity I and cross-section f falls on the layer. α, β are the glancing angles of the incident and scattered beam with respect to the surface of the powder, so that

$$\beta = 2\theta - \alpha \qquad \text{......(6)}.$$

To meet the actual experimental conditions we consider a beam diverging
a from the anticathode or from a slit A, figure 1, situated at a distance a from the centre of the layer. The scattered radiation is recorded on a photographic film in

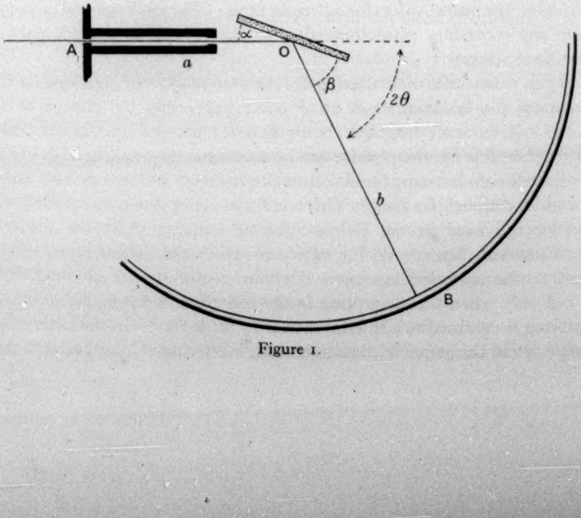


Figure 1.

B, placed on a circle with radius b, having O as centre, so that the X-rays fall normally on the photographic film. For this arrangement

$$P_{b\,(hkl)} = X\frac{p_{(hkl)}\, F^2_{(hkl)}\, N^2\, q\, \chi \sin\beta}{16\pi\, b \sin^2\theta\, \mu \cos\frac{1}{2}(\beta-\alpha)} \qquad \text{......(7)},$$

is the total intensity of the line representing the reflection (hkl) on the film, when a beam subtending the solid angle χ of intensity X per unit solid angle is incident on the layer.* The absorption coefficient enters into equation (7) in $1/\mu$ and is therefore a constant for all reflections from the particular layer. The quantities X, χ and b have common values for the exposure or are constants of the instrument and N is a constant for the particular type of crystal. The link between relative F values and observed P_b values is thus given by

$$P_{b\,(hkl)} = \text{const.}\frac{p_{(hkl)}\, F^2_{(hkl)}\, q_{(hkl)}}{\sin^2\theta}\, \frac{\sin\beta}{\cos\frac{1}{2}(\beta-\alpha)} \qquad \text{......(7$'$)}.$$

The values of p, q and θ are particular to each reflection and depend on crystallographic data of the lattice only, while the effect of the orientation of the layer with regard to the angles of incidence α and of scattering β is determined by the factor i where

$$i = \frac{\sin\beta}{\cos\frac{1}{2}(\beta-\alpha)} \qquad \text{......(8)},$$

which by introducing equation (6) can be written

$$i = \frac{\sin(2\theta-\alpha)}{\cos(\theta-\alpha)} \qquad \text{......(8$'$)}.$$

It will be seen that with other conditions equal the intensity for a given reflection, i.e. for a given angle θ, increases with β and with the difference $(\beta-\alpha)$. For small angles the intensities are approximately proportional to β.

§ 6. THE TECHNIQUE OF INTENSITY MEASUREMENTS WITH A FLAT LAYER

Expressions (7$'$), (8) and (8$'$) can be applied in different ways to the evaluation of intensities.

One way is to make use of a focusing condition of the flat layer, which is established for $\sin\alpha/\sin\beta = a/b$, when sharp lines are obtained on the film in B. A method based on this property in which the powder layer is rotated and a rotating screen is used in front of the photographic film, so as to satisfy the focusing condition over an extended angular range, has been described by the author in earlier papers$^{(2,8)}$. For this case we can write equation (7$'$) in the form

$$P_{b\,(hkl)} = X\frac{p_{(hkl)}\, F^2_{(hkl)}\, N^2 q_{(hkl)}\, \chi}{8\pi b \sin\theta}\, \frac{1}{\mu}\, \frac{1}{1+\sin\alpha/\sin\beta} \qquad \text{......(9)}.$$

---

* Expression (7) can be derived from

$$P_b = X\frac{pF^2 N^2 q\chi}{8\pi b \sin\theta}\, \frac{1}{\mu}\, \frac{1}{1+\sin\alpha/\sin\beta}$$

introduced in an earlier paper$^{(2)}$ by considering that

$$\frac{1}{1+\sin\alpha/\sin\beta} = \frac{\sin\beta}{2\sin\theta\cos\frac{1}{2}(\beta-\alpha)}.$$

The term $1/(1+\sin\alpha/\sin\beta)$ in (9) takes the place of the term i in the previous expression and is constant so long as the focusing condition is satisfied for a given ratio a/b. This term is thus eliminated in relative measurements.

To obtain the gain of intensity derived from an increase of angle β over α as given by equation (8), and at the same time satisfy the focusing condition, the distance a should be made smaller than b(9).*

The other way is to set the powder at a certain angle α with respect to the incident beam and to associate with each value of P_b the appropriate i term introducing in equation (8′) the angle θ belonging to each individual reflection. This method is particularly suitable for comparing lines situated within a small angular range. The angle α is then so chosen as to satisfy the focusing condition for an intermediate point of the range. This procedure has been adopted in the measurement on the dispersion of the scattering-power of tungsten. By choosing the focused angle to be nearer to one or the other line, the sharpness and density of one line can be increased as compared with another. This effect can be used within limits to adjust the densities of unequal lines.

The arrangement making use of a fixed angle α can also be used to compare lines comprising a wider angular range of θ outside the focused angle, and we have discussed this as applied to measurements on aluminium in a paper published in 1931(10). Measurements were then made comprising the same reflections, with the powder set at the fixed angles $\alpha=25{\cdot}0^\circ$ and $\alpha=35{\cdot}5^\circ$ in order to make possible an internal check by comparison of corresponding reflections in each case. Reference may be made to these exposures in order to indicate some points which arise in this connexion. Figure 2*a* is a reproduction of a film and of its microphotometric record† with the powder set so that $\alpha=25{\cdot}0^\circ$, figure 2*b* relates to an exposure with the powder set so that $\alpha=35{\cdot}5^\circ$. With $a=b$ the focused angle was 50° for the first and 71° for the second. Different reflections were therefore near the focused angle in each case and this accounts for the different broadening of the lines. For the purpose of examining more closely the sharper lines in both exposures, the relevant data are collected in table 1. Column 5 gives the relative i-values of corresponding lines for the two settings, to which the intensities of the reflections should be proportionate. When these are compared with the observed P_b values in column 6, expressed in similar relative units, the agreement is reasonably good, having regard to the quality of some of the lines. It will however readily be seen that the lines are not suitable for exact comparisons owing to the varying spread and intensity dis-

---

[*] It should be noted however that the focusing condition introduces a geometric-optical relation between the entrance slit in A and the line in B, so that for a given reflection a linear magnification between slit and line exists in the ratio $a:b$. By this relation the specific intensity of the reflections on to the film is reduced when b increases over a. This does not, however, apply when the actual width of the line is determined by factors other than the magnification of the slit, such as the penetration of the rays into the layer or the imperfection of the crystallites. Whether it is of advantage to make b larger than a depends thus on the particular conditions. In some of the measurements on tungsten a large distance b of the film from the powder was required in order to obtain sufficient separation of the lines, and the width of the lines was essentially determined by the broadening resulting from the smallness of the crystallites. Here it was found of advantage to make a smaller than b.

[†] The dense films chosen for better reproduction are not those actually evaluated by scattering(14).

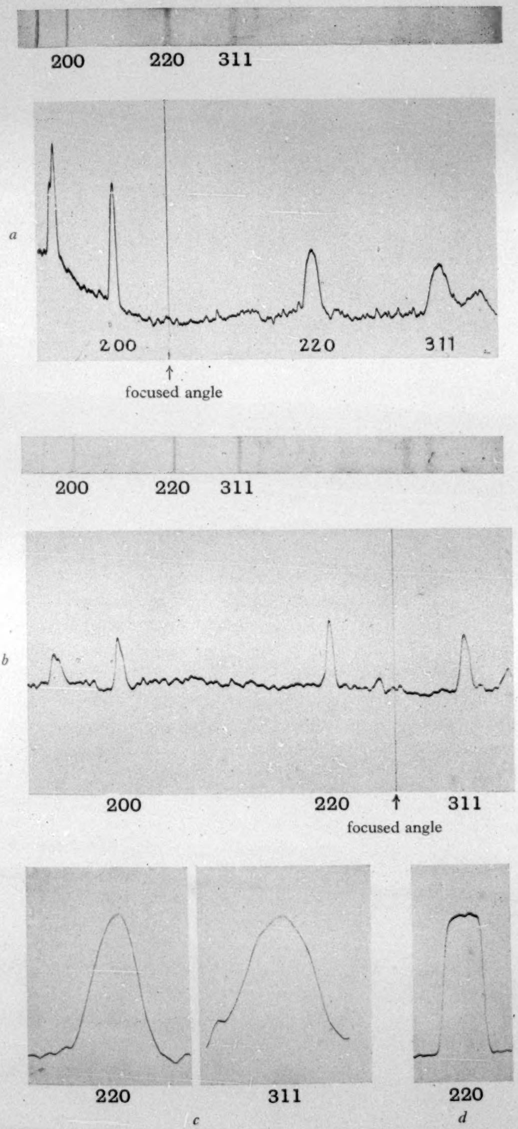


Figure 2. Scattering of Cu K_α radiation from stationary aluminium powder layer set at $\alpha = 25{\cdot}0°$, figure 2 *a* and at 35·5°, figure 2 *b*. ↑ denotes the focused angle. Comparative records of reflections taken outside the focused angle figure 2 *c*; in focus with wide slit figure 2 *d*.

tri
di
co
a
o

T

tribution, according to their distance from the focused angle. This is shown more distinctly by the photometric records, in which the ratios of the peak values differ considerably from the ratios of *i*-values in column 5, indicating that focusing plays a predominant part in determining the peak values and the density-distribution of the lines.

Table 1. Values of i or $\sin(2\theta-\alpha)/\cos(\theta-\alpha)$ for the settings $\alpha=25{\cdot}0^\circ$ and $\alpha=35{\cdot}5^\circ$ and various reflections from aluminium. Ratios of i and of P_b for the two settings (relative values).

Index	θ	i_{25°	$i_{35\cdot5^\circ}$	$\frac{i_{25^\circ}}{i_{35\cdot5^\circ}}$	$\frac{P_{b25^\circ}}{P_{b35\cdot5^\circ}}$
200	22·2°	0·333	0·159	2·10	2·10
220	32·4°	0·646	0·490	1·32	1·39
311	38·8°	0·818	0·673	1·21	1·10

§ 7. THE EFFECT OF FOCUSING

The distribution of the density across the line has a considerable bearing on the accuracy of the measurement when photographic methods are employed. The accuracy decreases when large and small densities have to be photometered side by side and it is desirable that the lines to be compared should have similar densities. Errors do not arise so much in the evaluation of levels of different density, represented by steps of adequate area when a blackening curve can be introduced with a high degree of certainty, as in the evaluation of an area of varying density when the density of each element has to be associated point to point with the appropriate intensity-value. This is particularly felt when relevant parts of the line extend through a range of widely varying densities. It is therefore desirable to obtain lines with a central part of approximately uniform density and a steep falling off on both sides to the density of the background. With crystallites giving sufficient definition an approach to this type of line can be obtained with a flat powder layer by satisfying approximately the focusing condition and using a wide entrance slit in *A*; the line will then be an image of this slit. Figure 2*d* is the microphotometric record of the aluminium (220) reflection obtained in this way, with an entrance slit 0·8 mm. wide. Figure 2*c* is a microphotometric record at equal magnification of the (220) and (311) reflections as in figure 2*a*. It will be seen that in this case the gradual decrease of density is not suitable for accurate measurement. It is for this reason that when we are recording a group of lines simultaneously with one fixed setting α of the layer we prefer to embrace only a limited angular range of about 20° for small angles and of 50° for large angles of deflection, and to use the rotating screen for larger angular ranges, when high accuracy is required.

Brindley and Spiers(3) recently described and discussed the use of a flat layer for quantitative intensity-measurements without having apparently noticed our earlier work on the subject. They deduce an intensity expression similar to our term (8′) and apply this to a layer set at a fixed angle α, observing the lines for different reflections in the same way as we have done in the measurements on aluminium to which we referred. They also discuss again the method of powder mixtures, which we described and used throughout in our earlier work.

It is very welcome to see the method of flat powders discussed from various points of view and the paper of Brindley and Spiers has the special merit of giving prominence to the effect of the angular term, which they call the absorption term, in determining the relative intensities of reflections at various angles. They indicate in a graph the change of it with angle α and refer to this as a means for controlling the relative densities of lines situated at different angles. It should be noted however for practical application that the i term affects the total intensity of the line as represented by P_b, while the varying broadening of the lines according to their position with regard to the focused angle (which is a distinctly different effect) is the major factor in determining the peak values and the density-distribution of the lines. We differ from Brindley and Spiers in considering that this broadening is not favourable for good measurements when it leads to considerably enlarged lines. The broadening outside the focused angle is reduced and i becomes the deciding factor when very narrow beams are used. This however reduces not only the intensities, but also the area of powder used for scattering, which impairs the averaging and is a disadvantage, particularly when the powder is held stationary as suggested by Brindley and Spiers.

With very small particles or with particles constituted of light atoms, when the X-rays penetrate to a greater depth and a greater number contribute to scattering, a stationary layer can be used; and such a layer can also be used with comparative immunity for large angles of deflection θ, when the penetration in the layer is again increased and a greater number of particles contribute to scattering. We have shown(8) that the broadening of a line outside the focused angle is proportional to $\epsilon\eta \cot \beta\, a/(a+b)$, where ϵ is the angular width of the incident beam in the plane AOB and $\eta/2$ the angular distance of the reflection from the focused angle. This broadening effect due to departure from the focusing condition is thus very much reduced for large angles and lines situated over a much wider range can be compared under conditions of approximate focusing. Measurements at large angles are generally undertaken under much more favourable conditions. Their use is limited because of superposition effects due to the crowding of many reflections when mixtures of different substances are used, so that it is difficult to determine the correct background. In many cases also the evaluation of F-values for small angles is of particular interest.

§ 8. THE EFFECT OF THE SIZE OF CRYSTALLITES

Another factor which affects the density-distribution depends on the limited resolving power of small or distorted crystallites comprising only a limited number of regularly spaced lattice planes. Broadening of the lines becomes noticeable when the number of regularly spaced planes is less than 1000 and increases with decreasing number. The intensity-distribution of the lines depends then on the size-distribution of the scattering elements. Side by side with the condition that crystallites must not be too large, imposed by extinction, we have thus the requirement that the crystallites should not be too small. Particularly unfavourable conditions are found for a powder containing a mixture of larger and smaller crystallites extending into the range of very small units, which we described as a vanishing size-distribution; this has been discussed in another paper(10). But apart from this case a powder consisting of small crystallites gives an unfavourable falling off of intensity which is undesirable for accurate evaluations.

This is not the place to discuss the preparation of suitable powders in detail; it may be recalled from previous papers(1, 2, 10), however, that the process of mechanical grinding is generally inadequate: the particles obtained are too coarse, at the same time they embody, particularly in the case of ductile metals, a certain number of highly deformed crystallites presenting all the characteristics of a vanishing size-distribution. Chemical precipitation and controlled processes of chemical reduction have proved successful in a number of instances, and here it was often found useful to introduce a spacing material to avoid the formation of larger units. This was, for instance, done in the measurements by Baxter and the author on tungsten. Four or five volumes of silica had been mixed with one volume of tungsten trioxide powder prior to reduction. This method has been proved effective to avoid the formation of larger particles.

The angular range through which a certain reflection is scattered in so far as it is due to the limitation of the scattering crystallites can be taken as inversely proportional to the number of regularly spaced lattice planes*, while with crystallites extending uniformly in all directions their number in unit volume is inversely proportional to the third power of the linear dimensions. With powder particles oriented at random the number of particles which contribute to the scattering for any reflection for a given setting of the powder layer and a given aperture of the incident beam is thus inversely proportional to the fourth power of the linear size of the crystallites. This accounts for the rapid variation of this number with the size of the scattering units. The actual numbers depend on experimental conditions, and in particular on the volume from which scattering takes place and on the angular aperture of the beams. For average conditions the number of crystallites taking part in the scattering of a narrow incident beam with crystallites of 10^{-6} cm. is of the order of 10^7 to 10^8, which is more than adequate to give good averages. For a size of 10^{-5} cm. the number of scattering crystallites would be 10^3 to 10^4, while larger crystallites give distinct unevenness of the reflected intensities.

In order to improve the averaging, when reflections are recorded simultaneously without rotating screen and with the powder set at a certain angle α, we give a small rocking motion to the layer extending through approximately 0·2° to 0·6° about an axis vertical to the plane *AOB*, figure 1. This is particularly effective in reducing the unevenness due to the presence of a limited number of larger units. Such crystallites even when they represent only a small fraction of the total mass and bring no appreciable extinction error, produce accidental inequalities which affect the photometric evaluation. For such units the range of reflection as determined from the dynamical theory of scattering is of the order of 5″. Thus a rocking motion of 30′ increases the number of reflecting elements with regard to any particular direction of incidence about 300 times. The rocking should be limited to a sufficiently small range, so as not to affect the intensity ratios as given by (8) and (8′). For large values of α and β the change of this expression with α for a small displacement to either side of the prefixed value is linear, so that the effects of symmetrical displacements to both sides cancel out. For small values of α and β this is not the case and the range of rocking should be made small accordingly. The treatment given to the deviations from the correct setting in an earlier paper(2) can be applied to this case.

---

* This refers to an arrangement with an approximately linear source.

§ 9. THE METHOD OF MIXED POWDERS

When the absolute value of F is being determined from the reflections of the powder layer we have to refer to expression (7) in the case of a flat layer or to a similar expression for other cases which require side by side with the measurement of the scattered intensity P_b the measurement of the incident beam and of the coefficient of absorption. It is difficult to introduce accurate numerical values for these quantities.

To meet these conditions we have resorted to the method of the quantitative powder mixture, which consists of using a powder layer containing a mixture of two kinds of crystals in the mass ratio m_1/m_2(2). If both types of particles are so small as to conform with the conditions of a uniform layer as discussed in §§ 3, 4 the same absorption coefficient applies to the reflections from each type of crystallites and we obtain P_b-values for each of them related to a common intensity of the incident beam and to a common absorption coefficient μ. The quantities X and μ do not therefore enter into the expression for relative values and we have for the ratio

$$\frac{P_{b(hkl)_1}}{P_{b(h'k'l')_2}} = \frac{m_1\,\rho_2\,p_{(hkl)}\,F^2_{(hkl)_1}\,N_1^2\,q_1\sin^2\theta_2}{m_2\,\rho_1\,p_{(h'k'l')}\,F^2_{(h'k'l')_2}\,N_2^2\,q_2\sin^2\theta_1}\cdot\frac{i_{(hkl)}}{i_{(h'k'l')}} \qquad \ldots\ldots(10),$$

where the suffixes 1 and 2 refer to the two kinds of particles and the suffixes (hkl) and $(h'k'l')$ mark two particular reflections.

The densities ρ_1 and ρ_2 of the two types of particles appear in this expression, as in equation (4), to account for the fact that the ratio of the masses per unit volume enter into P_b.

By this method the measurement of the ratio of the scattered intensities interlinks the values F_1 and F_2. An unknown F can be related to a value of F which is known from measurements on single crystals or for which calculated values are available.

§ 10. ABSORPTION EFFECTS IN COARSE POWDERS

The introduction of expression (10) implies that a common absorption term applies to both types of crystals. This is satisfied when the absorption in the individual particle produces a negligible reduction of intensity of the transmitted

X-ray beam and when the particles are sufficiently well mixed to conform with the requirement indicated in discussing the absorption coefficient of a powder in § 3.

These points have been overlooked in some cases and Schäfer(11) refers to the error which arises when the method of powder mixtures is applied to larger particles. He refers to this as accounting for discrepancies which appeared between measurements by Glocker and Schäfer(12) and the results of subsequent calculations by Hönl(13). In discussing the error introduced by absorption he considers in the first place the effect of defective balancing of the absorption in a column of particles of a mixture containing a limited number of elements.

A different treatment may be given here, which considers in the first place the absorption in the individual particle and is applicable in the most direct way to the conditions of a powder layer.

In considering particles which possess appreciable absorption two distinct cases have to be distinguished. In the first case the particle consists of one single crystallite. In this case in general (but with the exception of very heavy atoms like uranium and very long X-ray wave-lengths of 2 A. or more) it can be shown by introducing numerical values that primary extinction begins to affect the measurements for a particle-size smaller than that for which absorption becomes appreciable. The requirement of negligible extinction thus sets a limit to the particle-size at which absorption can be neglected, so that no consideration need be given to an absorption effect in the individual particle. In the second case the individual crystallites are sufficiently small to conform to the condition that extinction and absorption be negligible, but the particles are formed of larger aggregates which may possess appreciable absorption. This is the case which needs more detailed consideration.

We assume, as the simplest model of a layer, particles of two microcrystalline substances 1 and 2 forming cubes of equal size arranged in strata parallel to the surface of the powder, as indicated by figure 3. Let z be the side of the cubes and μ_1 and μ_2 the absorption coefficients for the two types of particles. If we consider radiation incident and scattered in directions near to the normal to the surface so as to leave out oblique transitions from one particle into adjacent particles of different type, the scattering for the first stratum is given by the integration of equation (2) for a flat layer of thickness z. Thus we obtain

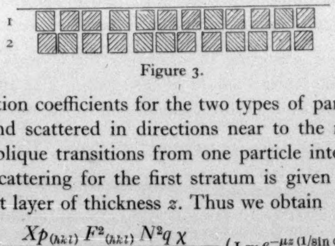


Figure 3.

$$P_b = \frac{X p_{(hkl)} F^2_{(hkl)} N^2 q \chi}{8\pi b \sin\theta \, \mu (1 + \sin\alpha/\sin\beta)} \left(1 - e^{-\mu z (1/\sin\alpha + 1/\sin\beta)}\right) \quad \text{......(11)},$$

and for the ratio P_{b_1}/P_{b_2} of two reflexions (hkl) and $(h'k'l')$, changing as from (9) to (7)

$$\left(\frac{P_{b(hkl)_1}}{P_{b(h'k'l')_2}}\right)_{\mu_1\mu_2} = \frac{m_1 \rho_2 p_{(hkl)} F^2_{(hkl)_1} N_1^2 q_1 \sin^2\theta_2}{m_2 \rho_1 p_{(h'k'l')} F^2_{(h'k'l')_2} N_2^2 q_2 \sin^2\theta_1} \frac{\mu_2}{\mu_1} \frac{\sin\beta_1}{\sin\beta_2} \frac{\cos\frac{1}{2}(\beta_2 - \alpha_2)}{\cos\frac{1}{2}(\beta_1 - \alpha_1)} \cdot \frac{1 - e^{-\mu_1 z (1/\sin\alpha_1 + 1/\sin\beta_1)}}{1 - e^{-\mu_2 z (1/\sin\alpha_2 + 1/\sin\beta_2)}} \quad \text{......(12)}.$$

This value has to be compared with the value $(P_{b_1}/P_{b_2})_\mu$ scattered from an intimately mixed layer, to which the coefficient of absorption as given by equation (4) applies, and containing the two constituents 1 and 2 in the same mass ratio. We assume the same scattering volume in each case and do not consider the absorption in the interspaces, which is not relevant.

R

We obtain a measure for the absorption error R by comparing these ratios

$$R=\left(\frac{P_{b(hkl)_1}}{P_{b(h'k'l')_2}}\right)_{\mu_1\mu_2}\Big/\left(\frac{P_{b(hkl)_1}}{P_{b(h'k'l')_2}}\right)_{\mu} \quad \text{......(13).}$$

Introducing the above values this becomes

$$R=\frac{\dfrac{\mu_2}{\mu_1}\cdot\dfrac{1-e^{-\mu_1 z(1/\sin\alpha_1+1/\sin\beta_1)}}{1-e^{-\mu_2 z(1/\sin\alpha_2+1/\sin\beta_2)}}}{\dfrac{1-e^{-\mu z(1/\sin\alpha_1+1/\sin\beta_1)}}{1-e^{-\mu z(1/\sin\alpha_2+1/\sin\beta_2)}}} \quad \text{......(13').}$$

R expresses how much the ratio of the intensities $(P_{b_1}/P_{b_2})_\mu$ scattered from the first stratum is modified by the fact that the two types of particles present different absorption compared with the ratio of these intensities if the same absorption coefficient applied to both. If $\mu_2>\mu_1$, the term R is the factor by which the intensity P_{b_2} should be multiplied to correct for the greater absorption loss.

The correction is definite when the directions of incidence and scattering are normal to the layer. Considering this case more in detail when $\alpha=\beta=90°$ the expression simplifies thus

$$R=\frac{\mu_2}{\mu_1}\,\frac{1-e^{-2\mu_1 z}}{1-e^{-2\mu_2 z}} \quad \text{......(14).}$$

We see that this becomes unity when $\mu_2=\mu_1$.

When $e^{-2\mu_1 z}$ and $e^{-2\mu_2 z}$ are both small, i.e. when the absorption in the individual particle is large, expression (14) becomes μ_2/μ_1, which indicates the maximum error with coarse particles. When the absorption in both types of particles is small, i.e. when $e^{-2\mu_1 z}$ and $e^{-2\mu_2 z}$ are both nearly unity, the limiting value of expression (14) becomes equal to unity, as we should expect.

Table 2. Term (14) for scattering in stratum I for various ratios of absorption coefficient μ_2/μ_1 and various values of absorption. The absorption is expressed by the intensity loss $1-e^{-\mu_2 z}$ for transmission in the more absorbing particle.

$1-e^{-\mu_2 z}$ / μ_2/μ_1	0·05	0·1	0·2	0·3	0·4	0·5	0·7	0·9
1·5	1·019	1·034	1·072	1·113	1·158	1·208	1·318	1·442
2	1·022	1·052	1·111	1·178	1·249	1·334	1·540	1·818
4	1·047	1·078	1·170	1·281	1·410	1·560	2·405	2·760
10	1·058	1·103	1·223	1·350	1·520	1·727	2·350	3·725

Intermediate values are illustrated in table 2, in which are plotted values of expression (14) for different values μ_2/μ_1 and different values of $1-e^{-\mu_2 z}$, which quantity measures the absorption of radiation in passing through the particle with greater μ. With great disparity between μ_2 and μ_1 the error is appreciable even when the individual particles absorb little. Thus when $\mu_2/\mu_1 = 10$ with an intensity-loss of only 5 per cent in the more absorbing particle the error is 5·8 per cent. When $\mu_2/\mu_1 = 1{\cdot}5$ with a corresponding loss of 10 per cent the error is only 3·4 per cent.

Successive strata give decreasing contributions to the scattered intensity owing to the absorption in the strata in front. When the absorption in the front stratum is very large, the successive layers will not alter appreciably the intensity-ratios as scattered by the front stratum and the discrepancy as compared with the intensity-ratio obtained from a finely subdivided powder is given by the ratio μ_2/μ_1 of the absorption factors. Such a case is of no practical interest.

In considering successive strata with particles having little absorption in each individual stratum, we have to distinguish the case in which corresponding particles in adjacent layers are of the same type figure 3 (*a*), and that in which they are of different type, figure 3 (*b*). In the first case the absorption error introduced by the size of the particles will be increased owing to the exponential character of the absorption law. In carrying out any closer estimates it will be necessary to take account of the shape- and size-distribution of the particles of each kind and of their numbers. It can however be seen that for the scattering from a greater number of strata the increased absorption effect due to the superposition of a large number of particles of the same type decreases. The probabilities of definite distributions are determined by terms in which the number n of the elements enters as a factorial, while the absorption is determined by terms in which n enters as an exponential. With increasing n the factorial terms gain over the exponential terms, which means that irregularities in the grouping of the particles become of lesser significance in determining the intensities. When we consider the case of oblique incidence the effect of the incident or the scattered beam passing through adjacent blocks can be referred to the case of transition through successive strata. Any more detailed treatment depends on the introduction of definite assumptions as to the shape and size of the particles and as to their relative numbers for which no actual data are available. The schematic treatment shows however that the absorption error of large particles as given by table 2, for one stratum, offers a criterion for estimating with a certain margin of safety the magnitude of the possible absorption error for a thicker layer.*

The practical question as to how far particles consisting of large aggregates of crystallites can actually be used for intensity-measurements depends not only on the absorption effect we have discussed, but also on the extent to which the particles can be considered to be free of larger crystallites. We have observed not only that in the case of ductile metals like gold, silver and platinum does the tendency exist for microcrystals to unite to larger crystallites, but also that in the case of brittle metals with a high temperature of recrystallization like tungsten some larger crystals are found embodied in the microcrystalline powder when no steps are taken to avoid the formation of large particles; the same thing was found with binary compounds like rock-salt. In such cases the intensity-measurements are affected by extinction. We consider therefore that in general, so far as the presence of larger crystallites is not eliminated, extinction effects are a potential cause of error and that with aggregate particles it is not safe to rely on their subdivision into uniformly small micro-crystallites. The best way for avoiding these errors is to limit the maximum extension of the crystallites by the actual size of the particles to a value for which extinction is small. In this case, as has been indicated above, the absorption error in the individual particle is negligible.

---

* When the layer consists of particles which satisfy the condition of negligible absorption in the individual particle and which are intimately mixed, the chance must be considered of several particles of the same type being placed close together when we obtain what is equivalent to an aggregate. Again, it is not possible to evaluate the irregularities introduced by such groups without introducing specialized assumptions as to the distribution with regard to sizes and numbers of particles. We can, however, obtain some estimate of the differential absorption effect which may arise from such groupings, by referring to expression (14) and introducing into it a small multiple of the mean particle-size. Large chance irregularities are very rare and chance groupings resulting from particles at larger distances being placed in line for any particular direction can be neglected when the layer is rotated or rocked.

§ 11. ACCURACY OF INTENSITY-MEASUREMENTS WITH POWDERS

As to the accuracy obtainable with these methods, it must be considered that when reflections are compared by means of expressions depending on the intensity relation (8) the intensities depend on introducing the correct angular values. For large angles α and β the expressions are not sensitive to small angular deviations, so that the accuracy with which the angular setting can be measured scarcely affects the results; when one of the angles is small the limits of error within which the correct setting can be established and in particular the limits of trueness of the powder surface make themselves felt and the error is found to be about 2 per cent for glancing angles of 15°, but increases rapidly for smaller angles.

A certain intrinsic error is attached to the photographic photometry. This is mainly expressed by the deviations found when repeated measurements of the same group of lines taken under similar conditions are compared. The limits of error introduced from this source with measurements of suitable lines, i.e. lines representing a well defined intensity-step, showing a flattened peak with rapid fall of density to both sides, is about 1 per cent, which is mainly determined by the local inequalities of the photographic emulsion. When lines with a gradual decrease of density are measured the accuracy is less. With measurements of the reflections from platinum black, which gives broadened lines, an accuracy of 6 per cent was found and the accuracy found with lines situated at greater distance from the focused angle, like the broader lines in figure 2, is similar. The error depends then very much on the distribution of the density through the line. With lines corresponding to a vanishing size-distribution much larger errors are observed. Lines of this type are unsuitable for quantitative measurements.

REFERENCES

(1) Brentano, J. C. M. and Baxter, A. *Z. Phys.* **89**, 720 (1934).
(2) Brentano, J. C. M. *Phil. Mag.* **4**, 620 (1927); **6**, 178 (1928).

(3) Brindley, G. W. and Spiers, F. W. *Proc. phys. Soc.* **46**, 841 (1934).

(4) Darwin, C. G. *Phil. Mag.* **44**, 433 (1926). Bragg, W. L., Darwin, C. G. and James, R. W. *Phil. Mag.* **1**, 897 (1926).

(5) Ewald, P. *Handb. der Physik*, **23** (2), (1933).

(6) Claassen, A. *Phil. Mag.* **9**, 57 (1930).

(7) Rusterholz, A. *Helv. phys. Acta*, **4**, No. 2 (1931).

(8) Brentano, J. C. M. *Proc. phys. Soc.* **37**, 184 (1925).

(9) Brentano, J. C. M. and Adamson, J. *Phil. Mag.* **7**, 507 (1929).

(10) Brentano, J. C. M. *Z. Phys.* **70**, 74 (1931).

(11) Schäfer, K. *Z. Phys.* **86**, 738 (1933).

(12) Glocker, R. and Schäfer, K. *Z. Phys.* **73**, 289 (1933).

(13) Hönl, H. *Z. Phys.* **84**, 1 (1933); *Ann. Phys.*, Lpz., (5) **18**, 625 (1933).

(14) Brentano, J. C. M., Baxter, A. and Cotton, F. W. *Phil. Mag.* **17**, 370 (1934).

DISCUSSION

Messrs G. W. Brindley and F. W. Spiers. In reading this paper our attention has been directed to a previous paper by the author in the *Zeitschrift für Physik* **70** (1931) which had escaped our notice. In our previous paper(3) we discussed and gave references to the author's previous work and very much regret not having seen his 1931 paper in which the use of a flat stationary plate of powder was described; we offer him our sincere apologies. In putting forward our own paper we wished to emphasize particularly the method of controlling the intensities of X-ray lines by the variable focusing and absorption obtained with a flat layer of powder, a method to which the author raises certain objections in § 1 of his paper. In our own work we have found the method simple to use and reliable.

Author's reply. I regret if my note conveys the impression of a too strong criticism of the paper by Messrs Brindley and Spiers. Not only in the 1931 paper, but also in later work, as in that of Baxter and the author where use was made of a stationary or quasi-stationary layer, the lines were of necessity not equally in focus and the weaker lines were then placed nearer to the focused angle so that their peak values were enhanced. We found it better, however, to confine ourselves to a small angular range and to observe the lines near to the focused angle using at the same time a wide entrance slit, in preference to placing the lines far outside the focused range, where the change of the i factor becomes important enough to constitute actually a controlling factor of the intensity ratios. In the present paper, the object of which is to set forth the conditions for obtaining the best accuracy with powder methods, these various aspects had to be discussed. As has been indicated in the paper certain conditions are more suitable for greater angular ranges than others; they are found in particular with light atoms and large deflections, where the loss of accuracy in comparing lines at a great distance from the focused angle is less marked.

535.421

THE THEORY OF THE FORMATION OF AN IMAGE BY A PLANE BAND GRATING USED IN THE SOFT X-RAY REGION

BY F. F. P. BISACRE, O.B.E., M.A., B.Sc.

Received April 29, 1935

ABSTRACT. This paper describes an investigation of the reflection of light from a plane grating illuminated by a cylindrical wave diverging from a Huyghens line source. It is shown that for pencils sufficiently narrow the isophasics (properly parabolas) become circles with a virtual focus as centre. The position of this virtual focus is found. A formula is given for the intensity of the light along the reflected isophasic. It has a Fresnel integral as a factor. It is shown that if the dimensions of the grating are chosen so that the modulus of the Fresnel integral has its maximum value, the intensity along the isophasic varies after the fashion of an Airy image and the maximum possible intensity occurs on the axis of the reflected pencil—in other words, the reflected light is automatically focused. The critical (optimum) length of the grating for automatic focusing is determined by the condition that the quadratic term in the expansion for the optical path in powers of distance measured along the grating face from its centre must be three-eighths of a wave-length. Formulae for the dispersion and resolution of an optimum grating are given and the resolution turns out to be exactly the same as for an ordinary grating, namely the total number of lines on the grating multiplied by the spectral order. Some numerical examples are given. Third-order effects have been considered and it is shown that in the conditions contemplated in the use of these optimum gratings, the third-order term affects the length of the optical path by something like one 650th part of a wave-length, and is consequently negligible.

FOR some time past plane gratings have been successfully used for photographing soft X-ray spectra, from say 10 to 75 A., without lenses. These gratings are lightly ruled on glass and are used at grazing angles of incidence. Glass is chosen instead of a metal because of its high resistance to corrosion(1) in soft X radiation. The grating operates by means of the flats between the grooves and is lightly ruled to keep the width of each flat roughly half the pitch—the best proportion. It must be used at grazing angles to ensure a high reflecting-power(2). The gratings are ruled with a band a few hundred lines only, instead of the usual several thousand. The precise width of the band is a point of considerable importance. That such a grating if ruled with the right number of lines does possess remarkable automatic focusing power—much as a concave grating does—is proved by the excellent photographs of spectra reproduced in papers by Siegbahn and Magnusson(3) and J. A. Prins(4) for instance.

Prins gives a rough rule for determining the proper number of lines: The quadratic term in the expansion for the optical path in powers of distance measured along the grating face from its centre must not exceed half a wave-length. It may be of interest to examine the theory of this automatic focusing in some detail, in the hope of obtaining definite information as to (*a*) the optimum width of the band of lines; (*b*) the nature of the images formed—in particular the nature of the isophasics and the distribution of intensity along them; and (*c*) the positions of the centres of these images and consequently the validity of the usual formula giving the directions of the lines.

Siegbahn and Magnusson's apparatus(3) is sketched in figure 1. The whole apparatus is contained in a vacuum chamber which can be attached to a vacuum pump. The X-ray generating apparatus is on the left. The target is made of a chosen material, such as tungsten, and the electrical pressure used was some 5000 V. Two slits are provided, the first being an ordinary mechanical slit some 0·01 mm. wide while the second is formed by a wedge placed very close to the face of the grating, on the plane originally used by Seemann in his crystal grating work. The photographic plate is at *P* and is at right angles to the central axis of the slits. The back of the grating is not parallel to the front face on which the lines are ruled, so that light reflected from the back of the grating is thrown into what is usually a harmless direction. The bands of lines used by Siegbahn on different gratings ranged from 1 to 3 mm. He built a special ruling-machine and ruled a set of gratings with from 300 to 1800 lines per mm. and bands of 1 to 3 mm. The best results appeared to be got with a 1-mm. band having 1800 lines, a slit-width S_1 of about 0·01 mm. and an incidence grazing angle of about 1°. The paper does not state the wedge slit-width; glancing angles of incidence from 0·5° to 3° are mentioned.

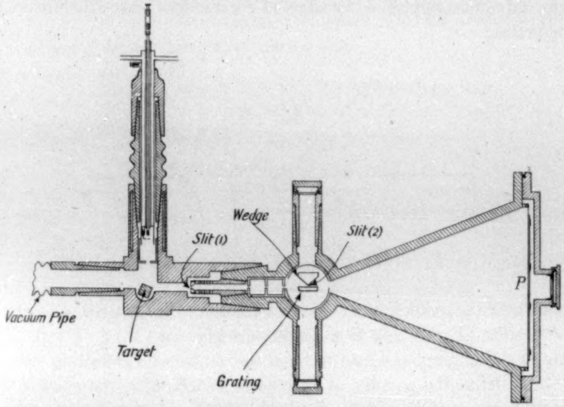


Figure 1. Vacuum plane-grating spectrometer*.

The use of two parallel slits with an extended source of radiation secures (*a*) a narrow band of the source as the illuminant; (*b*) narrow pencils of light from the elementary Huyghens sources constituting the active source-band; and (*c*) that the axes of these pencils shall be all very nearly coaxial with the centre-line of the slits.

---

* From Siegbahn and Magnusson, *Z. Phys.* **62**, 443 (1930), by the courtesy of Julius Springer, Berlin.

The angular aperture of the pencils is probably less than 1 minute of arc with the actual dimensions of the apparatus and slit-widths used. At all events, they are of this order.

We are thus led directly to a study of the screen image formed when a very narrow pencil of light from a Huyghens source is reflected from a plane grating. The actual image will be the sum of these elementary (non-coherent) images coming from the Huyghens sources forming the exposed active source-band, as defined by the pair of slits. We assume that the angular aperture of the incident Huyghens pencil, as controlled by the slits, is sufficiently wide to overlap the grating slightly, so that any diffraction ripples on the edges of the incident wave-front do not actually reach the grating.

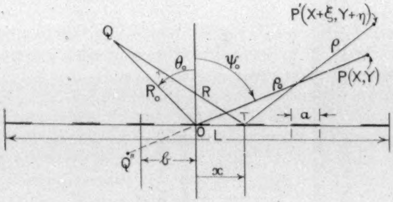


Figure 2.

We shall treat the problem as a 2-dimensional one. Let Q, figure 2, be a Huyghens source, P a point (X, Y), and P' a neighbouring point $(X+\xi, Y+\eta)$. Let O be the centre of the grating, assumed to be at the centre of a reflecting strip.

X, Y, ξ, η

R_0, θ_0

ρ_0, ψ_0

Let R_0, θ_0 define the position of Q in polar coordinates, the positive direction of the angle being as shown in the figure; and let ρ_0, ψ_0 define the position of P, again with ψ_0 positive in the direction of the arrow in the figure.

The coordinates ξ, η fix P' relative to a set of rectangular axes through P and parallel to the main rectangular axes, viz. the grating-face and the normal at O, as shown in the figure.

L The grating extends a distance $\frac{1}{2}L$ on each side of O.

x Let T be any point on one of the reflecting strips whose coordinates are x, 0.

R, ρ Let R be equal to QT and ρ to $P'T$.

E_z'

We suppose the light to be polarized in the plane of incidence, so that the incident electric vector E_z' is wholly along the Oz axis, i.e. it is perpendicular to the plane of the figure. It is given by

$$E_z' \approx \frac{B}{\pi} . e^{i\pi/4} . \frac{e^{-i\alpha R}}{\sqrt{R}} . e^{ipt} \qquad \ldots\ldots(1).$$

B where B is a constant, this being a convenient asymptotic form for the exact Bessel function representing a divergent cylindrical wave from a line source. The quantity

α, λ, p α is $2\pi/\lambda$ where λ is the wave-length, and p is 2π times the frequency of the light.

It can be shown that the value of the electric vector $E_{z,P'}$ of the reflected field at P' is given by the integral, taken over the reflecting surface of the grating,

$$E''_{z,P'} \approx -\frac{iB}{\pi\sqrt{\lambda}}\int \frac{e^{-i\alpha(R+\rho)}}{\sqrt{R\rho}}.\cos nR.ds.e^{ipt} \qquad \text{......(2).}$$ $E''_{z,P'}$

provided the reflecting power of the surface is unity. Consequently this is a suitable limiting form for both metals and glass if these are used under conditions where the reflecting-power is approximately unity. If this condition is not satisfied no such formula as equation (2) is available and the problem becomes much more complex. We can therefore take equation (2) as a guide, at all events, to what happens provided the reflecting surface has a reflecting-power sufficiently near to unity.

When $E''_{z,P'}$ has been determined, the whole reflected electromagnetic field follows at once. A simple calculation shows that

$$R = R_0 + x\sin\theta_0 + x^2.\frac{\cos^2\theta_0}{2R_0} + \text{higher terms in } x.$$

If the incident light were parallel the expression would end exactly with $x\sin\theta_0$.

Also
$$\rho^2 = (X+\xi-x)^2 + (Y+\eta)^2.$$

$$\therefore \quad \rho \approx \rho_0 + (\xi - x)\sin\psi_0 + \eta\cos\psi_0 + \frac{1}{2\rho_0}[(\xi-x)\cos\psi_0 - \eta\sin\psi_0]^2 + \ldots,$$

where
$$X = \rho_0\sin\psi_0 \text{ and } Y = \rho_0\cos\psi_0.$$

Hence
$$R+\rho \approx R_0 + \rho_0 + x(\sin\theta_0 - \sin\psi_0) + \xi\sin\psi_0 + \eta\cos\psi_0$$
$$+ \frac{1}{2\rho_0}[(\xi-x)\cos\psi_0 - \eta\sin\psi_0]^2 + \frac{1}{2R_0}x^2\cos^2\theta_0 + \ldots$$

as far as the second order in x, ξ and η.

Now $R+\rho$ occurs in our integral as

$$e^{-i\alpha(R+\rho)} = e^{-i(2\pi/\lambda)(R+\rho)},$$

so it is only the excess of $(R+\rho)$ over *any* multiple of λ that determines the value of this expression.

Let us put x equal to $nb+\bar{x}$, where n is a positive or negative integer, b is the grating-pitch and $|\bar{x}| < a/2$, where a is the width of a flat strip. n, b $|\bar{x}|$, a

Also put q equal to $(\sin\theta_0 - \sin\psi_0)$. Then q

$$qx = qnb + q\bar{x}$$

and, if
$$qb = r\lambda, \text{ where } r = \pm 1, 2, 3, \ldots$$ r

$$qx = (\text{a multiple of } \lambda) + q\bar{x}.$$

We can therefore replace qx in the expression for $R+\rho$ by $q\bar{x}$, where

$$-\frac{a}{2} < \bar{x} < \frac{a}{2},$$

and $\bar{x}$ is now comparable to λ whereas x itself is not.

But the condition $qb = r\lambda$ is the ordinary grating formula

$$b(\sin\theta_0 - \sin\psi_0) = r\lambda,\ r = \pm 1, 2, 3, \ldots.$$

The integral in equation (2) becomes, approximately,

$$I \approx \frac{\cos\theta_0}{\sqrt{(R_0\rho_0)}} e^{-i\alpha(R_0+\rho_0)} . \int e^{-i\alpha q\bar{x}} . e^{-i\alpha P} a\bar{x},$$

where

P

$$P = (\xi\sin\psi_0 + \eta\cos\psi_0) + \frac{1}{2\rho_0}[(\xi - x)\cos\psi_0 - \eta\sin\psi_0]^2 + \frac{1}{2R_0}\cos^2\theta_0 x^2 + \ldots$$

$$= (\xi\sin\psi_0 + \eta\cos\psi_0) + A\left[x - \frac{(\xi\cos\psi_0 - \eta\sin\psi_0)\cos\psi_0}{2A\rho_0}\right]^2$$

$$- \frac{\cos^2\psi_0(\xi\cos\psi_0 - \eta\sin\psi_0)^2}{4A\rho_0^2} + \frac{(\xi\cos\psi_0 - \eta\sin\psi_0)^2}{2\rho_0} + \ldots$$

A and

$$A = \frac{1}{2}\left\{\frac{\cos^2\psi_0}{\rho_0} + \frac{\cos^2\theta_0}{R_0}\right\} \qquad \ldots\ldots(3).$$

Now choose ξ and η so that

$$(\xi\sin\psi_0 + \eta\cos\psi_0) + \frac{(\xi\cos\psi_0 - \eta\sin\psi_0)^2}{2\rho_0}\left\{1 - \frac{\cos^2\psi_0}{2A\rho_0}\right\} = 0.$$

The term $\{1 - (\cos^2\psi_0/2A\rho_0)\}$ is $\frac{\rho_0\cos^2\theta_0}{R_0}\Big/\left(\cos^2\psi_0 + \frac{\rho_0}{R_0}\cos^2\theta_0\right)$ and the equation becomes

$$2(\xi\sin\psi_0 + \eta\cos\psi_0) + \frac{(\xi\cos\psi_0 - \eta\sin\psi_0)^2\cos^2\theta_0}{R_0\cos^2\psi_0 + \rho_0\cos^2\theta_0} = 0.$$

ξ', η', ψ_0 Now change the axes at P from ξ, η to ξ', η', where the $\angle\xi' P\xi = \psi_0$, figure 3.

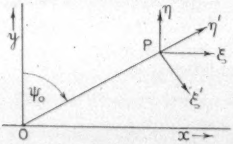


Figure 3.

Then

$$\xi' = \xi\cos\psi_0 - \eta\sin\psi_0,$$

$$\eta' = \xi\sin\psi_0 + \eta\cos\psi_0,$$

h so, if $\{R_0\cos^2\psi_0 + \rho_0\cos^2\theta_0\}/\cos^2\theta_0 = h$, the equation becomes

$$\xi'^2 + 2h\eta' = 0 \qquad \ldots\ldots(4),$$

which is a parabola, concave towards O.

So far as the radius of curvature at P is concerned this is indistinguishable from the circle

$$\xi'^2 + \eta'^2 + 2h\eta' = 0.$$

The centre of curvature is therefore at Q'' ($\xi' = 0$, $\eta' = -h$) i.e., is on PO produced, where

$$OQ'' = R_0\cos^2\psi_0/\cos^2\theta_0 \qquad \ldots\ldots(5).$$

Along the locus defined by equation (4) we have, then, for P

$$P = A\left[x - \frac{\xi' \cos\psi_0}{2A\rho_0}\right]^2$$
$$= At^2, \text{ say.}$$

Hence $$\int e^{-i\alpha q\bar{x}} \,.\, e^{-i\alpha P}\,dx = \int_{-a/2}^{+a/2} \sum_{n=-m}^{+m} .\, e^{-i\alpha(q\bar{x}+At^2)}\,d\bar{x},$$

where $$t^2 = \left[(nb+\bar{x}) - \frac{\xi' \cos\psi_0}{2A\rho_0}\right]^2.$$

Since n enters into t^2, only, we can put the integral into the form

$$\int_{-a/2}^{+a/2} e^{-i\alpha q\bar{x}} .\left\{\sum_{n=-m}^{+m} .\, e^{-i\alpha A t^2}\right\} d\bar{x}.$$

The sum is the sum of the values of the function

$$e^{-i\alpha A[X-C]^2},$$

where $$C = \xi' \cos\psi_0/2A\rho_0,$$

at $$X = \bar{x} - mb,\ \bar{x} - \overline{m-1}b,\ \ldots\ \bar{x} + mb,$$

all at distances b apart.

These are close enough (as examination in detail shows) to allow us to replace the sum by the integral,

$$\left.\begin{aligned}
\frac{1}{b}\int_{\bar{x}-mb}^{\bar{x}+mb} e^{-i\alpha A[X-C]^2}\,dX &= \frac{1}{b}\cdot\sqrt{\left(\frac{\pi}{2\alpha A}\right)}\cdot\int_j^k e^{-i\frac{\pi}{2}u^2}\,du \\
\text{where}\quad j &= \sqrt{\left(\frac{2\alpha A}{\pi}\right)}\,(\bar{x} - mb - C) \\
k &= \sqrt{\left(\frac{2\alpha A}{\pi}\right)}\,(\bar{x} + mb - C) \\
\text{and}\quad C &= \xi' \cos\psi_0/2A\rho_0
\end{aligned}\right\} \quad \ldots\ldots(6).$$

We can therefore put the sum as

$$\frac{1}{b}\sqrt{\left(\frac{\pi}{2\alpha A}\right)} . F(\bar{x}), \text{ where}$$

$F(\bar{x})$ is a limited Fresnel integral from j to k and involving $\bar{x}$ through its limits.

Consequently we have

$$E_{z,P'}{}'' \approx -\frac{iB}{\pi\sqrt{\lambda}}\cdot\frac{\cos\theta_0}{\sqrt{(R_0\rho_0)}}\,e^{-i\alpha(R_0+\rho_0)}\cdot\frac{1}{b}\cdot\sqrt{\frac{\pi}{2\alpha A}}\cdot\int_{-a/2}^{+a/2} e^{-i\alpha q\bar{x}} . F(\bar{x})\,d\bar{x} \quad \ldots(7).$$

This result of course includes all diffraction effects due to the limitation in the range of the Fresnel integral. It is only in quite exceptional cases that the presence of $\bar{x}$ in j and k affects the value of the Fresnel integral appreciably, for $\bar{x}$ is a very small quantity comparable to λ, whereas j and k are comparable to the grating-length.

As a rule, then, $F(\bar{x})$ can be taken outside the integration sign.

We then have
$$\int_{-a/2}^{+a/2} e^{-i\alpha q\bar{x}}\,d\bar{x} = \frac{a\sin\left(\frac{\pi ra}{b}\right)}{\left(\frac{\pi ra}{b}\right)},$$

so that
$$E_{z,P'}{}'' \approx -\frac{iB}{\pi\sqrt{\lambda}}\cdot\frac{\cos\theta_0}{\sqrt{(R_0\rho_0)}}\cdot e^{-i\alpha(R_0+\rho_0)}\cdot\sqrt{\frac{\pi}{2\alpha A}}\cdot\frac{a}{b}\cdot F\cdot\frac{\sin\frac{\pi ra}{b}}{\frac{\pi ra}{b}},$$

$$\begin{array}{lll} & \text{where} & \left.\begin{array}{l} F = \int_j^k e^{-i\frac{\pi}{2}u^2}\,du \\ j = -\sqrt{\frac{2\alpha A}{\pi}}\left(\frac{L}{2}+C\right) \\ k = +\sqrt{\frac{2\alpha A}{\pi}}\left(\frac{L}{2}-C\right) \\ C = \xi'\cos\psi_0/2A\rho_0 \end{array}\right\} \qquad \ldots\ldots(8). \\ j & & \\ k & & \\ C & \text{and} & \end{array}$$

This formula shows that the phase of $E_{z,P'}{}''$ is constant, except in so far as it is modified by the single factor F, the appropriate Fresnel integral.

We have localized P' so that it lies on the parabola through P, defined by equation (4) and having its centre of curvature at Q'', whose distance from O is $R_0\cos^2\psi_0/\cos^2\theta_0$, by equation (5).

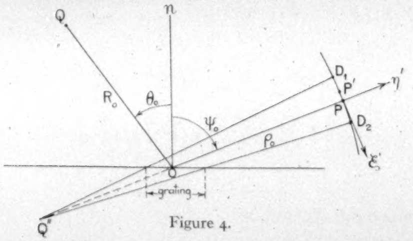


Figure 4.

This parabola is therefore an isophasic in the same sense as that in which a straight line is an isophasic for a fragment of a plane wave; it is an isophasic except for phase-ripples coming from varying values of the Fresnel integral, i.e. except for the diffraction that accompanies every wave-fragment at its edges. We have thus determined the basic isophasic for the reflected field, the central point of which is fixed by the grating formula

$$b(\sin\theta_0 - \sin\psi_0) = r\lambda, \quad r = \pm 1, 2, 3 \qquad \ldots\ldots(8a).$$

Special case. The field at P itself can be easily got. At P, $\xi' = \eta' = 0$, hence $C = 0$ and the range of F is from

$$-\sqrt{\frac{2\alpha A}{\pi}}\cdot\frac{L}{2} \text{ to } +\sqrt{\frac{2\alpha A}{\pi}}\cdot\frac{L}{2} \qquad \ldots\ldots(8b).$$

If each of these limits is sufficiently large to enable us to replace them, in the Fresnel integral, by $\mp\infty$, we get $F=\sqrt{2}.e^{-i\pi/4}$ and

$$E_{z,P}''\approx-\frac{Be^{i\pi/4}}{\pi}\cdot\frac{\cos\theta_0}{\sqrt{(R_0\cos^2\psi_0+\rho_0\cos^2\theta_0)}}\cdot\frac{a}{b}\cdot\frac{\sin\frac{\pi ra}{b}}{\frac{\pi ra}{b}}\cdot e^{-i\alpha(R_0+\rho_0)}\quad\text{......(9),}$$

the time factor being omitted.

When $r=0,\ \psi_0=\theta_0$ and if $a/b=1$,

i.e., when the grating is unruled, we get at once

$$E_{z,P}''\approx-\frac{Be^{i\pi/4}}{\pi}\cdot\frac{1}{\sqrt{(R_0+\rho_0)}}\,e^{-i\alpha(R_0+\rho_0)}\,.\,e^{ipt}\quad\text{......(10),}$$

a result agreeing exactly with equation (1). The source is now at the reflected image, Q' in the mirror, and there is a reversal of phase of π at reflection. Our formula (8) is therefore confirmed in this simple limiting case.

This result, at the centre of the image, applies only if the mirror is long enough to give us asymptotic Fresnel integrals.

To draw the isophasic, we find ψ_0 from the formula

$$b(\sin\theta_0-\sin\psi_0)=r\lambda,\ r=\pm1,\ 2,\ 3,\ \dots\quad\text{......(11)}$$

and set off OP, figure 4, equal to the given value ρ_0.

Produce PO backwards to Q'' where

$$OQ''=R_0\cos^2\psi_0/\cos^2\theta_0.$$

Q'' is the virtual image. Rule lines $Q''D_1$ and $Q''D_2$ through the ends of the grating, cutting the parabola defined by equation (4), in D_1 and D_2. Then D_1PD_2 is the fragment of the isophasic that interests us and we propose now to discuss the diffraction effects at the edges of this fragment.

Now $k-j=\sqrt{(2\alpha A/\pi)}\,L$ hence as we give C in equation (8) an increasing value, we merely decrease the upper limit and the lower limit. We therefore throw the arc of the Cornu spiral (of constant length $k-j$) towards one or other of the asymptotic points. If L is sufficiently large for $(k-j)/2$ to give approximately the asymptotic value of the integral at the centre of the image ($C=0$), then as we proceed along the isophasic from the central point P there will be little change in modulus or phase until j (or k) is zero, i.e., from equation (8),

$$\xi'=A\rho_0L\sec\psi_0\quad\text{......(12).}$$

when the amplitude will be halved.

Now the length of the arc D_1D_2, figure 4, cut off by the rays from Q'' to the ends of the grating is easily seen to be given by

$$D_1D_2\approx2A\rho_0\sec\psi_0\,.\,L\quad\text{......(13);}$$

hence the breadth of the image (defined as the width between half-maximum values of the amplitude) is the geometrical image.

We shall thus get the usual rectangular image flanked by ripples of phase and amplitude as sketched in figure 5, and the film will record a broad band of width D_1D_2.

These remarks apply when the image is received on a film normal to the beam. If the film is perpendicular to the plane of the grating, then the pencil is cut obliquely by the film and this obliquity may have to be allowed for. This image shows no focusing property at all. The light is merely spread out over a band which is the projection of the grating from the virtual focus Q'' on to the film.

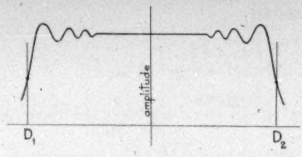


Figure 5.

Optimum (or focusing) conditions. Suppose now that instead of the upper limit of the integral $\int_0^d e^{-\frac{1}{2}\pi u^2}\,du$ being large enough to give asymptotic values, it is chosen just large enough to give the maximum modular value of this integral (not its asymptotic value). The upper limit will now have quite a small and perfectly definite value, viz., $d=\sqrt{\frac{3}{2}}$*, and

$$\int_{-d}^{+d} e^{-i\frac{1}{2}\pi u^2}\,du = 2\times 0{\cdot}945\, e^{-i\pi/4}$$

instead of

$$\int_{-\infty}^{+\infty} e^{-i\frac{1}{2}\pi u^2}\,du = 2\times 0{\cdot}707\, e^{-i\pi/4},$$

so that the central amplitude is now increased by 34 per cent and the intensity by 80 per cent.

If then we limit the length of the grating to secure this condition, we shall materially increase the central intensity of the image and modify the whole structure of the image; in fact, as we shall now show, we get an Airy type of image, instead of the rectangular pattern.

To secure this critical optimum condition, we get at once from equation (8*b*),

$$\sqrt{\frac{2\alpha A}{\pi}}\cdot\frac{L}{2}=\sqrt{\frac{3}{2}}$$

or, since

$$\alpha = 2\pi/\lambda,$$

$$A.\left(\frac{L}{2}\right)^2=\frac{3}{8}\lambda \qquad \text{......(13}a\text{)},$$

where

$$A=\frac{1}{2}\left(\frac{\cos^2\theta_0}{R_0}+\frac{\cos^2\psi_0}{\rho_0}\right)$$

= the coefficient of the second-order term in the expansion of $(R+\rho)$ for the central image point P, figure 2, in powers of x.

---

* This is not *exactly* the value of a but it is within 1 per cent of the true value and is used in this form for arithmetical convenience.

Hence $A.(L/2)^2$ is the maximum value of this second-order term and we get the rule: The extreme value of the second-order term, in the expansion of $R+\rho$ in powers of x, measured along the grating face from its centre, should be exactly three-eighths of the wave-length. If it is either more or less than this value, the central intensity and sharpness of the image will be impaired. Prins has given this critical value as $\lambda/2$, which appears to be rather on the high side.

In figure 6 the dark line shows the range of integration for the point at the centre of the image. As we move from the centre of the image outwards the arc of integration AB moves in one or other direction, remaining constant in length. Suppose it moves in the direction of the arrows in figure 6.

Then, clearly from figure 6, oscillations of modulus and phase in F begin immediately, so that changes of amplitude and phase in the image begin at once. In other words, diffraction effects begin at once.

The effect is to give an amplitude-band as illustrated roughly in figure 7, while the phase-ripple begins, very feebly, at the centre of the image. This type of image clearly does show a focusing effect. The range of integration is $2\sqrt{\frac{3}{2}}$.

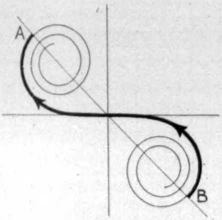


Figure 6.

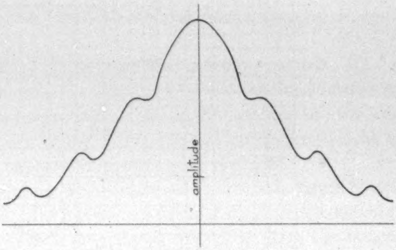


Figure 7.

The amplitude will, then, be proportional to y, where

$$y=\left|\int_{x-2\surd(3/2)}^{x} e^{-i\frac{1}{2}\pi u^2}\,du\right| \qquad \text{......(14)},$$

where $x=\sqrt{\frac{3}{2}}$ at the centre of the band.

Differentiating the integral in equation (14) with respect to x and equating to zero, we get turning-values of y (but not all of them), where

$$x=\sqrt{\tfrac{3}{2}}+\sqrt{\tfrac{2}{3}}.n,\quad n=1,\ 2,\ 3,\ \ldots \qquad \text{......(15)}.$$

The first occurs at $n=0$, i.e., at $x=\sqrt{\frac{3}{2}}$, the central line of the image.

Since $x_n=2\sqrt{\dfrac{A}{\lambda}}.\left(\dfrac{L}{2}-C_n\right)$, we get

$$\xi_n'=\pm A\rho_0\sec\psi_0\sqrt{\frac{2\lambda}{3A}}.n \qquad \text{......(16)},$$

But if L is the optimum length,

$$AL^2/\lambda=\tfrac{3}{2},\ \text{by } (8b) \qquad \text{......(16}a\text{)},$$

$$\therefore\ \xi_n'=\pm\rho_0\lambda n/L\cos\psi_0 \qquad \text{......(17)}.$$

V

If we denote by $2V$ the angle subtended by the grating at the centre of the image,

$$L \cos \psi_0/\rho_0 = 2V,$$

$$\therefore \quad \xi_1' = \lambda/2V \qquad \text{......(17a)},$$

which is precisely the same as the half-width of the Airy image formed by a lens of angular aperture $2V$.

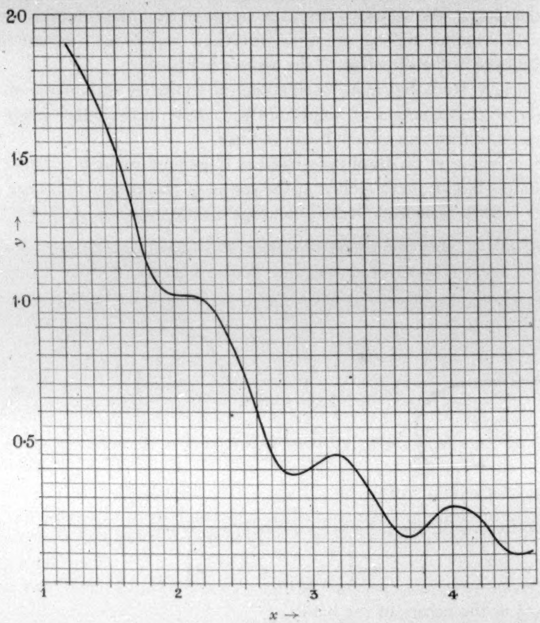


Figure 8.

$$y = \left| \int_{x-2\cdot 45}^{x} e^{-i\frac{\pi}{2}u^2}\, du \right| = f(x),$$

where $x = \sqrt{\tfrac{3}{2}} \left[1 - \xi' \dfrac{\cos \psi_0}{A \rho_0 L} \right]$

and θ_0 is angle of incidence;
ψ_0 is angle of diffraction;

L, length of grating;
R_0, distance of source from grating;

$$A = \frac{1}{2} \left\{ \frac{\cos^2 \theta_0}{R_0} + \frac{\cos^2 \psi_0}{\rho_0} \right\};$$

ρ_0, distance of centre of image from grating;
ξ', co-ordinate of field point.

A graph of y as a function of x, plotted from equation (14) is given in figure 8.

It will be seen that the turning value given by equation (15) for $n = 1$ is a maximum, with one minimum very close on the left of it. There is almost an inflection at

$$x = \sqrt{\tfrac{3}{2}} + \sqrt{\tfrac{2}{3}}, \text{ i.e. } n = 1.$$

The values of x for $n=2, 3, 4$, are all definite minima with maxima following them. There is no turning value for a value of x between $\sqrt{\frac{3}{2}}+\sqrt{\frac{2}{3}}$ and $\sqrt{\frac{3}{2}}$, so the central Airy band is of almost exactly the same width as in the ordinary telescope theory, viz. λ/V.

Phase-swinging. When $x=\sqrt{\frac{3}{2}}$ the phase of the Fresnel integral is $-42^\circ\, 33'$; when $x=\sqrt{\frac{3}{2}}+\sqrt{\frac{2}{3}}$ it is $-22^\circ\, 49'$, so that in passing across the central band from the centre to the minimum there is a phase-swing of about 20°, corresponding to $\frac{1}{18}$ of a wave-length. The parabola defined by equation (4) is thus not strictly an isophasic. The effect of this is that the true isophasic over the first band is not the parabola but a curve very close to it and differing from it at the first inflection by $\lambda/18$.

Now under the conditions in which these gratings are used at grazing angles, the actual half-breadth of the central image is some 6000 to 7000 wave-lengths. For example, if $\psi_0=87^\circ$, $L=0{\cdot}725$ mm. and $\rho_0=10''$,

$$2V=L\cos\psi_0/\rho_0=0{\cdot}000147,$$

and hence

$$\xi_1=\lambda/0{\cdot}000147=6800\lambda.$$

Consequently the parabola and the true isophasic are very nearly identical, in spite of the phase-swing.

The Poynting vector at any point of a (true) isophasic is orthogonal to it. If then we suppose the parabola to be the true isophasic, the error involved is that of neglecting the cosine of the angle between the normals to the parabola and the true isophasic at the point in question, as a factor. Since the two curves are so close to each other, the error in taking this cosine is obviously quite negligible.

We are therefore justified in squaring the ordinates of the curve in figure 8, and taking these numbers as proportional to the intensity of the field at different values of x.

Resolution. A very simple calculation now shows that if we place the images so that the maximum of one falls on the first minimum of the other, the combined intensity between the two maxima falls to 82 per cent of either maximum. From figure 8 we see that the values of y (the amplitude) are given by:

$$y=1{\cdot}01 \text{ when } x=2{\cdot}04,$$
$$y=1{\cdot}89 \text{ when } x=1{\cdot}225,$$
$$y=1{\cdot}37 \text{ when } x=1{\cdot}63.$$

The combined intensity-maxima are $(1{\cdot}89^2+1^2)$, which $=4{\cdot}57$; and the combined intensity-minima are $2\times 1{\cdot}37^2$, which $=3{\cdot}76=82$ per cent. This value corresponds to the 81 per cent of the Rayleigh rule. Consequently the conditions are almost exactly the same as those defined by the Rayleigh rule, and the breadth of the central band is the same as that of the Airy band; consequently we must have the same, the ordinary, resolution limit, Mr; in fact

$$-b\cos\psi_0\,.\,\delta\psi_0=r\delta\lambda_0,$$

i.e.,

$$b\cos\psi_0\,.\,\frac{\lambda}{L\cos\psi_0}=r\delta\lambda_0, \text{ by equations (17) and (17}a\text{)},$$

i.e.,

$$\lambda_0/\delta\lambda_0=rM \qquad \text{......(18)}, \qquad M$$

where M is the total number of lines on the grating.

It is rather remarkable and unexpected that the central band formed by such a grating should have the same width as the usual Airy band, and in spite of not being zero at the edge of the band should give the same resolution formula. The explanations are as follows. (1) Both bands are diffraction bands and hence depend on Fresnel integrals, in spite of the fact that the Airy image theory is most easily given as a first-order theory. (2) Since intensity is not zero at the first minimum, the combined maxima are considerably increased. Spectroscopists* do not anticipate so high a resolving power as Mr. It must be carefully borne in mind that theory only anticipates it provided the grating-length is chosen to comply exactly with the condition that the range of the Fresnel integral, for the centre of the image, shall be *exactly* $2\sqrt{\tfrac{3}{2}}$.

Minimal conditions. Just as there is an optimum length so there is a worst length—in fact, it would be just as easy to design the grating-length to give a minimum at the centre and brighter bands on either side, an image that might be mistaken for a double line. The conditions for this are obvious from figure 6. The choice of the proper length for optimum conditions depends on equation (16*a*), and to secure this either the grating must be ruled with the proper number of lines or the two slits must be chosen so as to pass the required very narrow pencil of light only.

Dispersion.

$$\frac{d\psi}{d\lambda} = \mp \frac{r}{b \cos \psi_0} \qquad \text{......(19),}$$

the upper or lower sign being taken according as ψ_0 is less than or greater than θ_0. The dispersion is unusually high when ψ_0 is nearly 90°, as it is in actual cases, but it still falls far short of the dispersion of a crystal grating.

Numerical examples. The figures for three cases are given in table 1.

The second column in table 1 is calculated for a grating of moderate pitch (1/15000″), an angle of incidence θ_0 equal to 89°, and a wave-length λ equal to 20 A. The grating is very short, only 0·725 mm. in length, and the number of lines is 427. The third column shows what happens if this grating is used at $\lambda = 40$ A., in which case of course the grating is not of optimum length. The central intensity is almost the same; the width of the central band increases by 43 per cent but the resolution is the same as before. The dispersion falls off by 27 per cent. The structure of the band does not alter radically over the range $\lambda = 20$ to 40 A. Consequently a grating designed for optimum conditions at $\lambda = 30$ A. might be expected to work satisfactorily between $\lambda = 10$ A. and $\lambda = 50$ A. The fourth column is calculated for a very fine pitch ($p = 1/45000''$). Comparing this column with the second, we see the glancing angle of reflection is nearly doubled; the angular aperture of the pencil is much as before; the length of the grating is 0·445 mm., 61 per cent of its former value; the number of lines goes up by 84 per cent; the central band hardly changes. But the dispersion goes up by 90 per cent and the resolution by about the same amount.

Overlap of images. The intensity-bands we have discussed arise from a Huyghens

---

* For example, Siegbahn (3), p. 2.

Table 1

Pitch, b	16,950 A. or 1/15000″	16,950 A. or 1/15000″	5650 A. or 1/45000″	Data
Order, r	1	1	1	—
Dimensions, R_0 and ρ_0 (in.)	10	10	10	—
Grazing-incidence angle, γ_0 (deg.)	1	1	1	—
Wave-length, λ (A.)	20	40	20	—
Grazing-reflection angle, δ_0	2° 57′	4° 4′	4° 56′	By Eqn. (8*a*)
Angular aperture, u (sec. of arc)	30	40	31	,, (17*a*)
Length, L (mm.)	0·725	0·725	0·445	,, (3) and (16*a*)
Number of lines, M	427	427	785	—
Width, $2\xi_1'$, of central band (A.)	$27{\cdot}3 \times 10^4$	$39{\cdot}2 \times 10^4$	$26{\cdot}5 \times 10^4$	,, (17)
Resolution, $\lambda/\delta\lambda$	427	427	785	,, (18)
Dispersion per angstrom, $d\psi/d\lambda$	$1{\cdot}145 \times 10^{-3}$	$0{\cdot}831 \times 10^{-3}$	$2{\cdot}06 \times 10^{-3}$	,, (19)
Range of Fresnel integral	$\pm\sqrt{\frac{3}{2}}$ or $\pm 1{\cdot}225$	$\pm 1{\cdot}16$	$\pm 1{\cdot}225$	—
Relative intensity	1	0·99	1	—

source on the extended source of radiation. The slits must be chosen so that the exposed line of the radiant source is sufficiently narrow to prevent serious displacement of the images formed by the elementary Huyghens sources of which the source line consists. It would be easy to destroy the whole advantage of this type of band by permitting too wide a line on the source to be exposed simultaneously. The exact conditions required will be evident from the calculations already given. The extreme Huyghens sources should be near enough to ensure that their respective images shall not be relatively displaced on the film by more than say 5 per cent of the central band-width. It may be impossible to secure this; but anything more will clearly lead to the production of a more diffuse band on the film.

Validity of the fundamental grating formula. It has been suggested that wave-lengths determined from the formula

$$b\,(\sin\theta_0 - \sin\psi_0) = \pm r\lambda$$

by plane-grating methods in the soft X-ray region are not quite correct. We have found no evidence for this view. We have shown that the centre of the image lies exactly in the direction given by this rule and it is the angular position of the centre of the image that it is the object of experiment to determine.

Third-order effects. Our theory includes only second-order terms in the expansion of $(R+\rho)$ but we can readily show that, in the conditions contemplated in a_3 table 1, the third-order term is negligible. The third-order coefficient is a_3, where

$$a_3=\frac{1}{2}\left[\frac{\sin\psi\cos^2\psi}{\rho_0^2}-\frac{\sin\theta\cos^2\theta}{R_0^2}\right] \qquad \ldots\ldots(20).$$

The maximum length that this term contributes to $(R+\rho)$ is a_3x^3, where x is the half-length of the grating or, in terms of the wave-length, $(a_3x^3/\lambda)\,\lambda$.

γ Now $b(\sin\theta-\sin\psi)=r\lambda$; hence, if $\theta=90^\circ-\gamma$

δ and $\psi=90^\circ-\delta$,

$$\tfrac{1}{2}b\,(\delta^2-\gamma^2)\approx r\lambda.$$

If $R_0\approx\rho_0$

$$a_3\approx\frac{1}{2\rho_0^2}[\delta^2-\gamma^2]$$

$$\approx r\lambda/b\rho_0^2$$

and

$$a_3x^3/\lambda\approx\frac{r}{b\rho_0^2}x^3$$

$$\approx\frac{r}{b\rho_0^2}\cdot\frac{L^3}{8}$$

$$\approx\frac{r}{8}\left(\frac{L}{\rho_0}\right)^2\frac{L}{b}$$

$$\approx\frac{1}{8}\left(\frac{L}{\rho_0}\right)^2 Mr,$$

where M is the total number of lines. This formula is true whether the grating is of optimum length or not, provided that $R_0\approx\rho_0$.

If $\rho_0=10''$, $L<1$ mm. and $Mr<800$, then

$$\frac{a_3x^3}{\lambda}<\frac{1}{645},$$

so that the *maximum* error in omitting this term is only some 645th part of a wave-length. It is difficult to see how the inclusion of this term could materially affect our results.

Note added July 20, 1935

In the July (1935) number of the *Proceedings* of this Society, Petrie$^{(5)}$ gives very interesting details of a single-slit plane spectrograph. In this apparatus a very obliquely inclined X-ray target is used as the source slit, and the width of the beam of light leaving the grating is limited by an adjustable slit placed very close to the end of the grating and at right angles to the grating-face. The grating itself is not ruled for optimum conditions—if it were, the beam leaving the grating would be already focused so that the slit would be superfluous. The image photographed with this instrument is the diffraction image of the slit used.

REFERENCES

(1) R. W. Wood and T. Lyman. "Improved gratings for vacuum spectrographs." *Phil. Mag.* p. 310 (July, 1926).

(2) H. M. O'Bryan. "Reflecting power and grating efficiency in the extreme ultra violet." *Phys. Rev.* **38**, 2nd series, 1, p. 32 (July, 1931).

(3) M. Siegbahn and T. Magnusson. "Zur Spektroskopie der ultraweichen Röntgenstrahlen, I." *Z. Phys.* **62**, Hefte 7 and 8 (1930).

(4) J. A. Prins. "Beiträge zur Plangitterspektroskopie der ultraweichen Röntgenstrahlen." *Z. Phys.* **69**, Hefte 9 and 10 (1931).

(5) D. P. R. Petrie. "Comparison of X-ray wave-lengths by the plane-grating vacuum spectrograph, and the structure of the *K* line of carbon." *Proc. phys. Soc.* **47**, 626 (1935).

621.318.423:621.317.334

A DYNAMOMETER NULL METHOD OF MEASURING THE INDUCTANCE AND THE EFFECTIVE RESISTANCE OF IRON-CORED CHOKES CARRYING DIRECT CURRENT

By GEOFFREY D. PEGLER, M.Sc.

Received March 8, 1935. *Read June* 21, 1935

ABSTRACT. An alternating potential-difference is applied to the terminals of two parallel circuits, one of which contains the fixed coils of a reflecting electrodynamometer, the inductance to be measured and a known variable resistance, while the other contains the moving coil of the dynamometer in series with a condenser of capacity K and a second known variable resistance. The former resistance is fixed during a determination, while the latter is varied until the phase-difference between the currents in the two branches is exactly $\frac{1}{2}\pi$, in which case there is no deflection of the moving coil. The inductance is then given by $L = KR_K R_L$, where R_K and R_L are the total effective resistances of the capacity and inductance branches respectively. A second balance is now obtained with the value of the first resistance altered by a suitable and known amount, the second resistance being altered accordingly. From the data of the two balances both the effective resistance (which is included in R_L) and the inductance may be calculated. The method is equally applicable when direct current is allowed to flow through the inductance. It is accurate over a range of inductances from a small fraction of a henry to that of the largest chokes. Examples of measurements are given.

---

§ 1. INTRODUCTION

In the discussion which followed a paper published in these *Proceedings*(1), a null or no-deflection method of using an electrodynamometer for measuring the inductance of iron-cored chokes was suggested. The present paper carries out this suggestion, and provides thereby an extremely simple means of accurately determining the inductance of a choke in terms of resistance and capacity, and also its simultaneous a.-c. resistance with or without the superposition of direct current.

The change of zero due to imperfect elasticity of the suspension, inherent in a deflectional method, is absent, and the absence of a deflection also dispenses with errors introduced by mutual inductance between the fixed and moving coils of the electrodynamometer. The balances necessary for a determination may be approached by the adjustment of a single variable. Since the frequency is not involved in the fundamental equation, the method is independent of the wave-form of the alternating current used, in so far as the quantities measured are themselves independent of frequency.

§ 2. THEORY

If alternating currents of the same frequency are flowing in both the fixed and moving coils of an electrodynamometer there will be a deflection of the moving coil unless the phase-difference between the two currents is exactly 90°, the sign of the deflection changing as the phase-difference passes through that value. The absence of a deflection signifies that the two currents are in exact quadrature and this fact may be applied to the measurement of inductance in the following way.

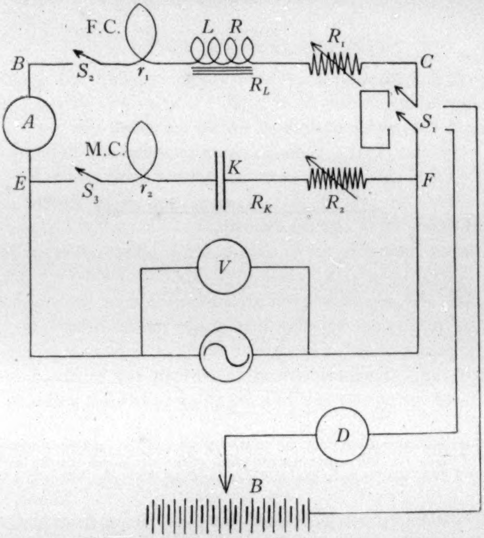


Figure 1 *a*. Diagram of circuits. *F.C.* is the fixed coil and *M.C.* the moving coil of the electrodynamometer; *LR* is a choke, *K* is a condenser, and R_1 and R_2 are variable resistance boxes. R_L and R_K denote the total resistances of the branches *BC* and *EF*. *A* and *D* are a.-c. and d.-c. milliammeters respectively, *V* is an a.-c. voltmeter; S_1, S_2, S_3, are switches, and *B* is an h.-t. battery.

An alternating voltage is applied, figure 1 *a*, to two circuits in parallel, one containing the fixed coils of the dynamometer in series with the choke of inductance L, to be measured, and a known variable resistance; the other circuit including the moving coil, in series with a condenser of capacity K and a second variable and known resistance. The current will lead the applied volts in the capacity branch by the phase angle ϕ', and will lag in the inductance branch by the angle ϕ''. Assuming simple harmonic currents and voltages, and denoting the total resistances in the inductance and capacity branches respectively by R_L and R_K and the r.m.s. values of the currents by I_L and I_K, we shall have figure 1 *b* as the vector diagram representing the phases of the currents and the applied voltage.

L
K
ϕ', ϕ''
R_L, R_K
I_L, I_K

When, by the variation of R_L or R_K, quadrature between I_L and I_K is obtained, we have

$$\tan\phi' \tan\phi'' = 1.$$

Since $\phi' = \tan^{-1} 1/pKR_K$, and $\phi'' = \tan^{-1} pL/R_L$, the condition for no deflection of the moving coil is therefore

$$pL/R_L \times 1/pKR_K = 1$$

or

$$L = KR_L R_K \qquad \text{......(1),}$$

a simple relation requiring no measurements of frequency, current, or applied voltage.

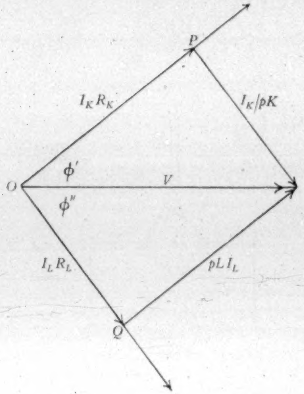


Figure 1*b*. Vector diagram representing the relative phases of the currents through the branches *BC* and *EF* of figure 1*a*. *OP* represents the phase of the condenser current leading the applied voltage *V* by ϕ', *OQ* the phase of the choke current lagging by ϕ'', while $\phi' + \phi'' = 90°$.

If, when the balance is being obtained, the phase defect from perfect quadrature
δ, k is δ, the torque on the moving coil may be written $kI_L I_K \sin\delta$, where k is an
instrumental constant. The differential coefficient of $\sin\delta$, has its maximum value
when δ is zero, with the desirable result that the instrument will be most sensitive
R_1, R_2 to phase-differences at the null point. If R_1 and R_2 denote the values of the re-
r_1, r_2 spective box resistances in the inductance and capacity branches, r_1 and r_2 the
R resistances of the fixed and moving coils of the electrodynamometer, R the a.-c.
R_0 resistance of the choke, and if further we denote the sum of R_1 and r_1 by R_0, we have

$$R_L = R_1 + r_1 + R = R_0 + R \text{ and } R_K = R_2 + r_2.$$

The formula in greater detail thus becomes:

$$L = KR_K (R_0 + R) \qquad \text{......(2).}$$

To find R, the value of R_0 is changed by altering R_1, and R_K is correspondingly
R_K', R_0' altered until a new balance is obtained. Denoting the new values by R_K' and R_0',
we have

$$L = KR_K' (R_0' + R) \qquad \text{......(2'),}$$

and the effective resistance is given by

$$R=(R_K R_0 - R_K' R_0')/(R_K' - R_K) \qquad \ldots\ldots(3).$$

R having been determined from equation (3), the inductance L may now be calculated by substituting this value in equation (2). The value of L so found includes the inductance of the fixed dynamometer coils, which is quite negligible when iron-cored chokes are being measured.

If the alternating voltage is supplied through a transformer with a secondary of low d.-c. resistance, direct current may be injected into the branch containing the choke without affecting the moving coil.

§ 3. EXPERIMENTAL CIRCUIT

To investigate the method experimentally, a reflecting electrodynamometer was constructed having two vertical parallel fixed coils of 500 turns each, of mean diameter 2·5 inches, and about 2·5 inches apart; and suspended centrally between these a moving coil of 300 turns, of mean diameter 1 inch, carrying a small mirror of about 1 metre radius. The number of turns in the coils is only limited by the heating effect, which must not be permitted to generate convection currents sufficient to disturb the zero of the suspended system. With the coils specified this effect was undetectable, while the accuracy of measurement secured was ample for the purpose, see table 1.

The suspension was a phosphor-bronze strip of length 3 in. and cross-section 0·007 × 0·0005 in., and a similar nearly-taut strip led the current out to the lower contact. The period of swing was about 6 sec., and a damping vane of mica attached below the coil rendered the movement nearly deadbeat. The resistance of the moving coil including the strip was 35 Ω., and that of the two fixed coils in series was 60 Ω.

K was taken from a calibrated, decade, 1-μF. condenser subdivided down to 0·001 μF., and R_1 and R_2 were taken from non-inductive dial resistance boxes giving 0 to 10,000 Ω.

The alternating voltage V was taken from the secondary of a transformer of low d.-c. resistance, the supply to the primary being regulated by a potential-divider, preferably non-inductive, connected across 220-volt 50-cycle mains.

By the aid of the double-pole throw-over switch S_1, figure 1 *a*, direct current can be injected into the choke circuit from a battery of small accumulators, this current being read on the d.-c. milliammeter D. The thermal milliammeter A registers the r.-m.-s. value of the total current. Alternatively it is possible, by means of a thermionic voltmeter, to read the alternating current component separately. Both the inductance and capacity circuits can be broken independently by means of the switches S_2 and S_3.

§ 4. EXPERIMENTAL PROCEDURE

The electrodynamometer is set up with the plane of its moving coil at right angles, as judged by eye, to that of the fixed coils. The terminals of the moving coil are short-circuited, while alternating current is passed through the fixed coils.

The suspension is adjusted until there is no change of zero of the moving coil when the current is switched on or off. The mutual inductance between the coils is now zero.

To test a choke with alternating current alone, the switch S_1 is closed to the left, V is set at a low value, S_2 is closed and S_3 left open to prevent needless deflection of the moving coil. V is increased until the alternating current at which the test is to be made is registered by the milliammeter A. S_3 is then closed momentarily and R_2 is adjusted, and this process is repeated until the position of the spot on the scale remains unaffected by the opening or closing of S_3. After the values of the box resistances R_1 and R_2 have been noted, R_1 is increased by a suitable amount and the foregoing procedure is repeated. From the data of these two successive balances both the inductance and the effective resistance of the choke may be calculated. Of course, the increase in R_1 should be of the same order as the effective resistance of the choke as found by a preliminary determination.

The switching necessary for a determination with direct current superposed is evident from the circuit diagram. The required direct current, registered by the d.-c. milliammeter D, is obtained by adjusting the wanderplug connexion to the battery, a small completing adjustment being made by varying R_1. V is altered until the r.-m.-s. sum of the required d.-c. and a.-c. currents is registered by A, S_3 is closed, and the procedure for two balances is followed as before, save that the value of the direct current must be kept unaltered by a readjustment of the battery connexion when R_1 is altered for the second balance. Preferably the a.-c. component is measured separately with a grid-leak type of thermionic voltmeter connected across a known resistance in the choke circuit.

If the currents through the choke are kept constant during a series of tests, and the values of R_0 and $1/KR_K$ for each balance point are plotted as ordinates and abscissae respectively, a straight-line graph intersecting the Y axis at $-R$ will result. The slope of the graph gives the value of L, so that the values of both L and R may quickly be obtained. They may also be calculated, of course, from the data by applying equations (3) and (2).

§ 5. EXPERIMENTAL RESULTS

Figures 2*a* and 2*b* represent a series of graphs plotted as described above for two chokes, and indicate the degree of dependence of their inductance and effective resistance on the value of the superposed direct current. The graphs in figure 2*a* refer to a choke of good design, rated at 20–30 henries, with a direct current-carrying capacity of 50 mA. and of which the d.-c. resistance was 387 Ω. The graphs in figure 2*b* appertain to a choke of inferior quality rated by its manufacturers at 20 henries, the d.-c. resistance in this case being 340 Ω. From figure 2*a* it will be observed that the effective resistance, as shown by the intercept on the vertical axis, varies slightly with the superposed direct current in the same way as the inductance, showing first an increase followed by a decrease. Both the inductance and the effective resistance of the second choke when no direct current is passing are seen to have very high values (117 H., 16,700 Ω.), which however fall rapidly when direct current is superposed. Thus when 12 mA. is flowing the values become 22·19 H. and 1400 Ω. The very large drop in the effective resistance shows how considerable may be the effect of a polarizing current on hysteresis and eddy current losses, to which the resistance is mainly due. A typical set of results, from which one of the graphs in figure 2*b* is drawn, is given in table 1.

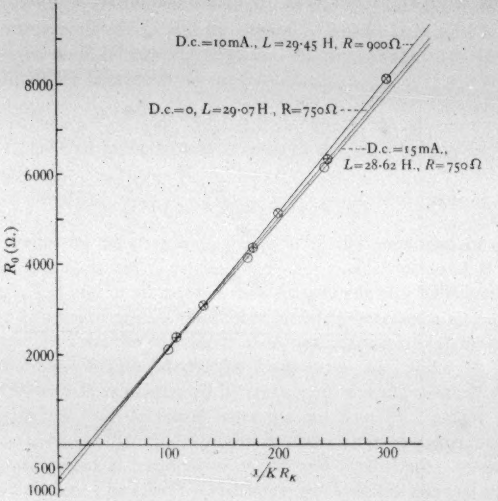


Figure 2*a*. Results for a choke of good design. Frequency of supply 50 c./sec. K was constant and equal to 1 μF. The a.-c. was in all cases 3 mA.

Table 1. $K = 1\,\mu$F., a.-c. = 3 mA., d.-c. = 8 mA., frequency = 50 c./sec., $r_1 = 60$ Ω., $r_2 = 35$ Ω.

V (V.)	R_1 (Ω.)	R_2 (Ω.)	R_0 (Ω.)	R_K (Ω.)	$1/KR_K$
31	1190	9570 ± 10	1250	9605	103·1
34	4280	4842 ± 3	4340	4877	204·9
39	6600	3550 ± 3	6660	3585	278·3
42	8150	3023 ± 2	8210	3058	322·0

Hence, from the graph, $R = 1900$ Ω. and $L = 30{\cdot}68$ H. when the superposed d.-c. is 8 mA. From column 3, showing the accuracy of setting, it is seen that the individual value of L can be regarded as correct to about 1 part in 1000.

To illustrate the applicability of the method to the measurement of small inductances, the choke was removed from the circuit and the inductance of the fixed coils of the electrodynamometer alone measured. In this case R_1 was kept constant, balances being made by giving K different values and varying R_2 until the deflection of the moving coil in each case was zero.

The graph connecting values of K with the corresponding values of $1/R_L R_K$ proved to be very closely linear and to pass through the origin. The value of L deduced was 0·049 H. ±1 per cent. A determination by a bridge method gave 0·0505 H.

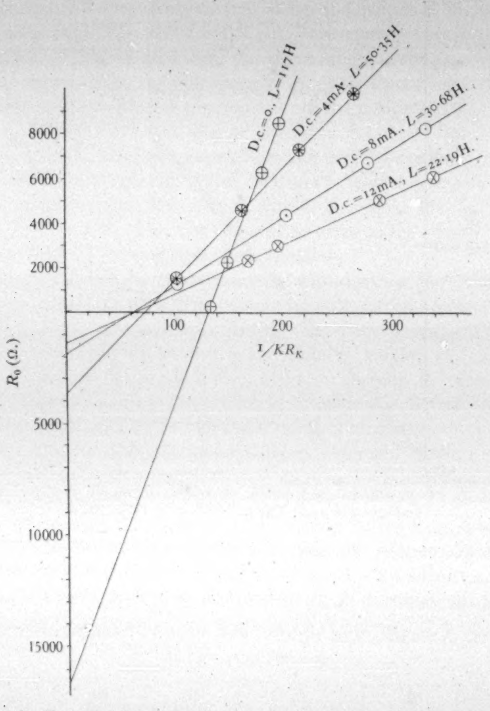


Figure 2*b*. Results illustrating the properties of a choke of inferior design. Frequency of supply 50 c./sec. The slope of the straight lines is a measure of the inductance; the intercept below the origin on the axis of ordinates is the effective resistance of the choke.

§ 6. RANGE AND ACCURACY

The method serves for a wide range of values of inductance. For chokes having very high impedance the voltage V must, of course, be correspondingly increased in order to provide the alternating current required for the test. To keep the current in the moving coil within suitable limits, the impedance of this branch must likewise be increased, and this may be effected by decreasing K and, at the same time, increasing R_K. Conversely, when small inductances are being measured V must be small, and to maintain high sensibility the impedance of the moving-coil branch must be correspondingly decreased by increasing K and decreasing R_K.

Errors may arise from two sources: (i) the inductance of the moving coil in the capacity branch, and (ii) the self-capacitance of the choke itself. The importance of both effects will depend on the frequency of the alternating current used, and on 50-cycle mains both effects will be negligibly small. In regard to (i) the inductance of the moving coil was 0·0032 H. and the largest value of K used was 1 μF. The exact formula for balance is

$$p\,(L+l_f)/R_L \times (1/pK - pl_m)/R_K = 1,$$

where l_f and l_m are the respective inductances of the fixed and moving coils of the electrodynamometer. l_f, l_m

Thus $$(L+l_f) = KR_KR_L/(1-p^2l_mK) \qquad \text{......(4)},$$

so that, at a frequency of 50 c./sec., the error in the measurement of L due to neglecting the term l_m is less than 1 in 3000. For smaller values of K the error is correspondingly less and in all cases at this frequency it may be neglected.

In making a measurement at the higher audio frequencies a knowledge of the frequency is required and the correction, as shown in equation (4), must be applied when the inductance is calculated. The linearity of the graphs drawn with coordinates $1/KR_K$ and R_0 is unaffected by frequency, and the effective resistance of the choke under test may be found from the graph or by equation (3) without correction. The choke for which the graphs in figure 2*a* were drawn, at a frequency of 50 c./sec., was tested at frequencies of 768 and 1000 c./sec., a valve oscillator being used as a source. The effective resistance was found to increase considerably, becoming 11,000 Ω. at 768 c./sec., and 18,000 Ω. at 1000 c./sec.

The large value of effective resistance of some chokes in comparison with their d.-c. resistance has been already remarked. It is clear that the frequently-employed method of calculating the impedance of a choke from its inductance and d.-c. resistance is unjustifiable.

In regard to (ii), if the self-capacitance of an inductance be regarded as a small condenser C in parallel with it, the effective value L' of the inductance will be given by $L' \doteqdot L/(1-p^2CL)$ and the effective resistance R' by $R' \doteqdot R/(1-2p^2CL)$, where L and R are its values at very low frequency.

In the measurement of inductance we are not, however, directly concerned with the self-capacity term, since it is the actual effective inductance and effective resistance under the known applied frequency that are required.

§ 7. ACKNOWLEDGMENTS

The above work was carried out in the Physics Research Laboratory of the Sir John Cass Institute, London. My best thanks are due to Dr D. Owen, upon whose suggestion the method is based, for his advice and interest throughout the work.

REFERENCE

(1) Willoughby. *Proc. phys. Soc.* **46**, 292 (1934).

DISCUSSION

Dr D. Owen. I believe the method will prove its value, serving as a quick and accurate means of measuring not only the self-inductance of large inductances, but also their effective resistance to alternating currents. The dynamometer is an instrument whose merits deserve the attention of engineers interested in accurate measurements, providing as it does an indicator applicable over the full range of audible frequencies and even beyond it. Its construction, suitable data of which for the measurements under present consideration are given by the author, offers no serious difficulty, nor is any special care needed in its adjustment for use.

It may be noted that, as the formula indicates, a null reading may be attained not only by varying the resistances, but also by varying the capacity. Now that accurate decade mica condensers are available at prices comparable with those of resistance boxes this consideration might be applied also to many bridge methods for the measurements of inductance or capacity. For instance, the use of a variable capacity eliminates the usual objection to Maxwell's inductance-capacity bridge, and should result in a return to that method, so is desirable otherwise on account of its fundamental simplicity.

Dr L. Hartshorn. In comparing methods for the testing of choke coils carrying both alternating and direct current, perhaps the most important consideration in practice is the possibility of making the test with the choke working at full load, i.e. carrying the largest possible direct current. It was this consideration which led me to choose Hay's bridge for such measurements some years ago, since this bridge satisfies the condition to a remarkable degree. The measurements mentioned in the paper are limited to chokes carrying very light loads, and some modification of procedure would probably be required for measurements under actual working conditions. It would be of interest to hear what are the capabilities of the method in this respect.

Mr A. Campbell. If the current in the condenser branch contains no harmonic components, then any such components as may occur in the L branch have no effect on the wattmeter reading; in this case the equation $L=KR_L R_K$ is quite valid, and the results obtained should be the same as if a method with a tuned detector were used. But if the supply voltage is not purely sinusoidal, in general both the branch currents will contain harmonics; and although a setting to give zero deflection can always be made, it seems impossible that the same equation should still hold. The author might find it interesting to explore such a case by testing ironless coils with a supply voltage of much-distorted wave form.

Author's reply. The only modification of the present method for the testing of chokes under full load conditions is that which was employed by Dr Hartshorn himself, namely the inclusion in the choke circuit of a fixed resistance specially wound to dissipate the heat generated at large values of direct current. In this case the two necessary balances would be made first with no other resistance than that of the choke itself in the inductance branch, and secondly with the fixed resistance included.

In the dynamometer method there is only one condition of balance to be satisfied, and this gives a distinct advantage over a bridge method with its two necessary balance conditions which are often not independent. The Hay bridge suffers from the latter disadvantage and also from the fact that its balance equations include the frequency, a fact of fundamental importance when the measurement of a choke with a core of variable permeability is to be made, since in this case the choke current will contain a high percentage of harmonics. Another advantage of the present method is that, when a balance point is being approached, the current need only be switched on for long enough to observe the direction of deflection of the moving coil, which indicates the sense in which the variable is to be altered. By this means the heating-effect in the resistance is reduced to a minimum.

I am obliged to Mr A. Campbell for pointing out a source of possible error due to harmonics in the supply voltage. It was shown in Mr Willoughby's paper, in which the deflectional method of using the electrodynamometer for such measurements was introduced, that to a first approximation the presence of harmonics cancelled out. In the null method, however, harmonics in the capacity branch will certainly tend to affect the accuracy of the observed readings, the amount depending on the permeability of the core, the value of the alternating current used, and the percentage of harmonic content in the supply voltage. I am indebted to Mr Campbell for his suggestion.

537.226

THE BREAKDOWN OF DIELECTRICS UNDER HIGH VOLTAGE, WITH PARTICULAR REFERENCE TO THERMAL INSTABILITY(1)

By S. WHITEHEAD and W. NETHERCOT

Received January 18, 1935. *Read in title June* 7, 1935

ABSTRACT. Provided a sufficiently wide range of external conditions can be applied, dielectrics exhibit both a thermal and non-thermal type of breakdown. The electric strength for thermal breakdown can be calculated from the stable electrical properties of the dielectric and a simplified theory is given in the paper which is shown to agree with experiment for a number of dielectrics for different conditions of temperature, time of application of stress, etc. A theory of ionization coefficients is developed to explain conductivity phenomena with a.-c. and d.-c. and ionization potentials are deduced. It is, however, shown that these phenomena do not appear to bear a direct relation to the possible types of non-thermal breakdown and that the ionization potentials have not the same physical significance as the corresponding quantities in gases. Experimental results relating to the type of non-thermal breakdown which occurred in the present investigations are given.

§ 1. INTRODUCTION

Origin of the investigation. The theory of the breakdown of dielectrics through thermal instability has already been studied fairly extensively(2), but mainly by means of breakdown tests in which only the field-strength, at which a marked and usually sudden instability occurs, has been noted and its variation with temperature, for example, has been compared with the corresponding variation of the conductivity. The absolute values of electric strength have often not agreed very closely with theory, while the materials examined and the ranges of temperature employed have been of a rather special character. In the work here described the behaviour of a number of well-known dielectrics was studied both before and during breakdown, at temperatures at which they are normally sufficiently stable to be employed for insulation purposes.

The main objects of the investigation were to determine whether the processes occurring before breakdown are compatible with a simple theory of thermal instability and whether the electric strength can be calculated from a knowledge of these processes in the stable range, even though other processes may occur at the instant of breakdown. In addition, it was desired to study the limiting features which determine the range of application of the theory with respect to temperature, time of stress and other factors.

General thermal instability: simplified theory. The energy-losses which occur in a dielectric under electric stress cause a temperature-rise distributed in a certain way. If the energy-losses increase with temperature and electric stress, a condition may occur when the generation of heat rises more rapidly with temperature than the thermal dissipation, and the temperature at some point or region will rise more and more rapidly, eventually causing a loss of insulating properties or an electric discharge. One of the simplest cases is that of a sheet of dielectric of thickness d in a uniform field with plane electrodes which cool the dielectric surface at a constant rate of λ watts per degree C. per cm.^{2} Suppose also that the energy-losses W at θ° C. above the ambient temperature are given by

$$W = E^2\, p_0\, e^{\gamma\theta} \text{ watts/cm.}^3 \qquad \ldots\ldots(1),$$

d

λ

W, θ

where p_0 is the specific energy-loss (a.-c. conductivity) at the ambient temperature, E the electric field and γ a constant. Utilizing Fock's$^{(3)}$ analysis, we find that the maximum voltage V which can be applied to the dielectric without causing thermal instability is as follows:

$$V = \sqrt{(2/p_0\gamma\rho)} \sin\alpha \exp\left[-\alpha\cos^2\alpha/(\alpha+\sin\alpha\cos\alpha)\right] \qquad \ldots\ldots(2),$$

p_0

E, γ

V

where ρ is the thermal resistivity of the dielectric in degrees C. per watt and α is given by

$$\lambda d\rho/2 = \sin\alpha \sec^3\alpha\,(\alpha+\sin\alpha\cos\alpha) \qquad \ldots\ldots(3).$$

ρ, α

We observe that $\lambda d\rho/2$ is P_i/P_e, where P_i is half the thermal resistance of the dielectric and P_e is the thermal resistance of an electrode. In other words, α is a function of P_i/P_e or the ratio of the internal to the external thermal resistance of each symmetrical half of the arrangement. We also observe that if the internal thermal resistance is negligible, then the maximum voltage V_m for stability is given by

$$V_m = \sqrt{(\lambda d/2\, p_0\gamma e)} \qquad \ldots\ldots(4),$$

whence

$$\frac{V}{V_m} = \sqrt{\left[\frac{2\exp\left(\frac{1}{2}\sin 2\alpha - \alpha\cos 2\alpha\right)}{\sec^2\alpha + (\alpha/\sin\alpha)\sec^3\alpha}\right]} \qquad \ldots\ldots(5).$$

P_i

P_e

V_m

Equation (5) shows that V/V_m, or the effect of the temperature-gradient in the dielectric, is a function of P_i/P_e. We may distinguish two extreme cases. When P_i/P_e is great the critical voltage V finally depends only on the electrical and thermal properties of the dielectric and is independent of its thickness, so that a certain maximum voltage is associated with a given dielectric for $\alpha \to \pi/2$ and $V \to \sqrt{(p_0\gamma\rho/2)}$. The calculation of this limiting voltage is usually fairly easy, but is difficult to verify experimentally owing to the high voltages involved and the intervention of extraneous effects, while in practice it is important only with very high-voltage equipment. Accordingly, if we consider the other extreme when P_i/P_e is small we observe from figure 1 that if $P_i/P_e < 0{\cdot}1$ the error which arises from neglecting the temperature-gradient in the dielectric will not exceed 1·5 per cent, while if $P_i = P_e$ the error is about 12 per cent. Although equation (5) is strictly true only when $W \propto E^2$ or the conductivity is independent of field-strength, yet the values when $P_e > P_i$ do not vary greatly if forms differing from equation (1) and of the more usual type occur. Neglecting P_i we then find the following simple equations for the limit of stability:

$$W = f(E_m, \theta), \quad E_m = V_m/d \qquad \ldots\ldots(6),$$

$$\partial W/\partial\theta = \lambda, \qquad \ldots\ldots(7),$$

E_m

θ being the temperature of the dielectric.

In the present paper, in addition to equation (1) which gives the solution (4), two general forms are considered:

$$\sigma = W/E_m{}^2 = p_0 + aE_m \exp\{(\gamma_0 + \gamma E_m)(\theta - \theta_0)\} \qquad \ldots\ldots(8.1),$$

and

$$\sigma = W/E_m{}^2 = p_0 + (a_0 + aE_m) \exp\{\gamma_0 (\theta - \theta_0)\} \qquad \ldots\ldots(8.2).$$

These lead to

$$\theta_a = \theta_1 \log_e (\theta_1/e\theta_E) + \theta_0 - \theta_p \qquad \ldots\ldots(9),$$

θ_a, θ_0, θ_1
θ_E, θ_p

where θ_a is the ambient temperature, θ_0 a constant, $\theta_1 = 1/(\gamma_0 + \gamma E_m)$ for (8.1) or $1/\gamma_0$ for (8.2), $\theta_E = daE_m{}^3/\lambda$ for (8.1) or $(a_0 + aE_m) E_m{}^2 d/\lambda$ for (8.2), $\theta_p = E_m{}^2 d p_0/\lambda$. Generally a sufficient approximation is obtained by writing

$$E_m{}^2 \sigma = W = p_0 E_m{}^3 e^{\gamma(\theta - \theta_0)} \qquad \ldots\ldots(10),$$

whence

$$E_m = \{\lambda/ed\gamma p_0\}^{\frac{1}{3}} \qquad \ldots\ldots(11).$$

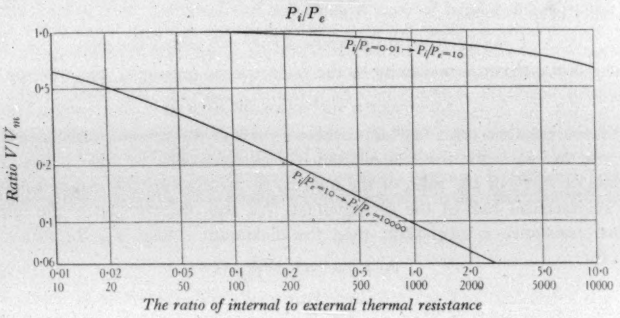


Figure 1. The effect of thermal resistance in the specimen on the thermal instability.

θ_c, t

V_m or E_m corresponds to the limit of thermal stability and therefore to the electric strength corresponding to a theoretically infinite time of application, since a slight disturbance would only cause the temperature to rise at an infinitesimal rate. If a greater electric stress is applied, the temperature will rise more rapidly and a critical temperature will be reached in a limited time. We might also consider that there is an intrinsic electric strength which decreases with temperature so that the temperature will rise until this electric strength is equal to the applied field. However, it is usually more in accordance with the facts to consider a limiting temperature θ_c at which, say, some chemical or physical change occurs. The time t for breakdown corresponding to a stress E greater than E_m is given by

$$\frac{\lambda t}{ds} = \sum_0^\infty \beta^{n+1} \left[\frac{n!}{(n+1)^{n+1}} - e^{-(n+1)\phi} \left\{ \frac{\phi^n}{n+1} + \frac{n\phi^{n-1}}{(n+1)^2} + \ldots + \frac{n!}{(n+1)^{n+1}} \right\} \right] \ldots(12),$$

, β, ϕ, W_{am}
W_a, θ_a

where s is the specific heat of the dielectric, $\beta = eW_{am}/W_a$, $\phi = \gamma\theta_c$, and W_{am} and W_a are the energy-losses at θ_a for field-strengths of E_m and E respectively. For equation (1), $\beta = e(E_m/E)^2$ and for equation (10) $\beta = e(E_m/E)^3$, so that $\lambda t/ds = f(E_m{}^2/E^2)$ and $f(E_m{}^3/E^3)$ in the two cases respectively. Thus the ratio of the electric strengths for two given times of breakdown is independent of the temperature at which the ratio is taken.

In actual fact, equation (12) may usually be simplified to

$$\lambda t/ds = \sum_{1}^{\infty} (n-1)!\,(\beta/n)^n - \beta e^{-\phi} \qquad \ldots\ldots(13).$$

If the ambient temperature is low compared with θ_c, a further simplification is possible, thus

$$\lambda t/ds = \beta\,(1 - e^{-\phi})$$

or

$$(E/E_m)^n = eds\,(1 - e^{-\gamma\theta_c})/\lambda t \qquad \ldots\ldots(14),$$

where $n=2$ or 3 according to the material. Equation (13) becomes, when θ_c is made very large

$$\lambda t/ds = \sum_{1}^{\infty} (n-1)!\,(\beta/n)^n \qquad \ldots\ldots(13.1).$$

This is identical with an equation given by Wagner⁽⁴⁾ for a rather different type of breakdown.

Experimental methods. The square law, or constancy of conductivity with respect to field-strength applies to few dielectrics at high field-strengths, and if moisture is present the departures are greater, while the effect of temperature is different at different field-strengths. Accordingly, for simplicity in interpretation of the results, it was necessary to ensure that the thermal resistance outside the specimen should be sufficiently great in comparison with that of the specimen. In addition, precautions had to be taken to ensure that concentrations of stresses and discharges in the ambient medium should not occur, while at the same time satisfactory contact had to be provided so that power factor and permittivity results should be valid, or at least repeatable. The first two of these conditions were fulfilled in one apparatus, figure 2 (*a*), by using rounded edges with oil immersion and a thermal backing of oak discs with cotton-wool packing; and in another apparatus, figure 2 (*b*), by using a very narrow gap between inner plate and guard ring and protecting the outer edge with mycalex, which is arc-resisting. A thermal backing of varnish-paper discs was used. The first apparatus was used in an oil* bath kept at uniform temperature, and the latter in an air thermostat. Contact with cellulose materials, such as ebonite, was made by painting graphite (aquadag) films on the surface; these give satisfactory contact under oil as well as in air. Plain brass electrodes were used for papers which, if flat, give repeatable values on the application of a suitable pressure, though these values cannot be taken as characteristic of the material. The external thermal resistance per specimen was 1·9° C./W. for figure 2 (*a*) and 3·2° C./W. for figure 2 (*b*)†. It was determined by replacing the specimens by flat eureka heaters made so that the energy-inputs under the inner plate and guard ring were proportional to their respective areas.

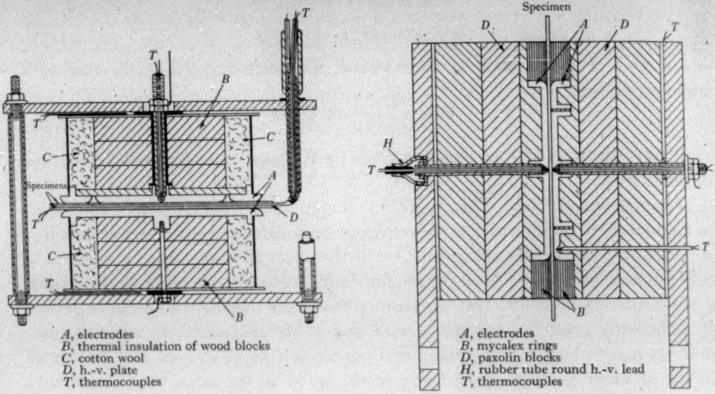
Such a test condenser formed one arm of a Schering bridge. A voltage was applied and maintained until stability of temperature, power factor and permittivity was approached; the voltage was then raised in successive steps until instability occurred with eventual breakdown, unless breakdown occurred first. A frequency of 50 c./sec. was used throughout.

---

* The oil, Wakefield's superforma, was colourless and the power factor was generally less than 0·001.

† Preliminary tests on varnished cloth were made with a simpler apparatus having a thermal resistance of 5° C./W.

Separate power-factor and permittivity measurements were made with similar electrodes, but with the thermal packing absent. D.-c. conductivity measurements were made under oil with electrodes similar to those shown in figure 2(*a*), but insulated with mica and glass and carefully screened to avoid leakage and stray charges. Short-time tests of electric strength were made with E.R.A. standard brass electrodes in view of the number of samples needed. The upper electrode is $1\frac{1}{2}$ in. in diameter with its edge rounded to a diameter of $\frac{1}{4}$ in., the lower electrode being a disc 3 in. in diameter. In such tests the time for breakdown is observed for a number of different field-strengths and the value for the given time interval is interpolated.


T
B
T
C
Specimen
A
D
T
C
T
B
A, electrodes
B, thermal insulation of wood blocks
C, cotton wool
D, h.-v. plate
T, thermocouples
Specimen
D
A
D
T
H
T
T
T
B
A, electrodes
B, mycalex rings
D, paxolin blocks
H, rubber tube round h.-v. lead
T, thermocouples

(*a*) Oil-immersed apparatus. (*b*) Apparatus for use in air thermostat.

Figure 2. Main forms of apparatus.

§ 2. PHENOMENA PRECEDING BREAKDOWN: ENERGY-LOSS FORMULAE

A.-c. conductivity phenomena. Cellulose acetate* and cellulose ethyl ether appear to have a species of ordered lattice structure, but behave, when they contain slight traces of moisture, as if two phases are present. The first is associated with a conductivity appreciable at lower temperatures but having a small temperature coefficient; the second has a negligible conductivity at lower temperatures, but a large positive temperature coefficient. This combination forms a well-known case of the Wagner-Maxwell theory of heterogeneous dielectrics and gives a total conductivity or specific energy-loss curve which has an inflection with respect to temperature as illustrated in figure 3*. At higher temperatures the conductivity is mainly determined by the second phase and increases linearly with electric stress and exponentially with temperature as shown in figure 4*, and in agreement with equations (8.1) and (8.2). The second phase, which gives the exponential term, is associated with the moisture present, as is shown by the decrease of this term as the moisture is progressively removed. This is illustrated in figures 5 and 6, the laws for which are similar to the equations quoted, namely

---

* The cellulose acetate contained a plasticizer such as tricresyl-phosphate or triphenol-phosphate, but from tests made with pure films (prepared by A. M. Thomas) the plasticizer appears to have only a minor effect on the electrical properties considered here.

$$\sigma = p_0 + aE \exp \gamma (\theta - \theta_0) \text{ for cellulose acetate} \qquad \ldots\ldots(15.1),$$

and

$$\sigma = p_0 + aE \exp (\gamma_0 - \gamma E)(\theta - \theta_0) \text{ for cellulose ethyl ether} \ldots\ldots(15.2).$$

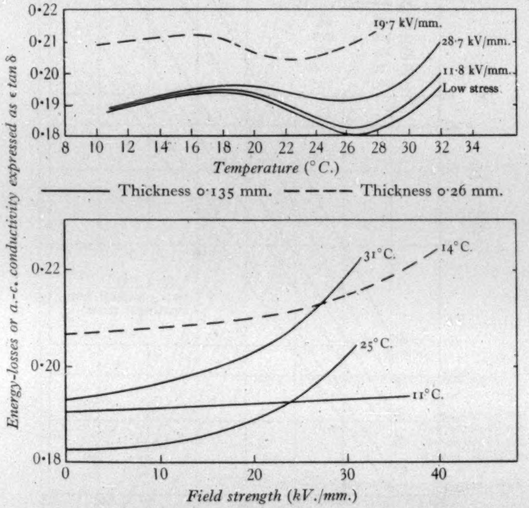


Figure 3. Effect of field-strength and temperature on cellulose-acetate films containing moisture at lower temperatures. (Conditioned 1 hour at 30° C.)

Approximately it may be said that with all these materials equation (10) was obeyed, the departures from this simpler form increasing as the moisture was removed.

A similar conclusion applies to thicker papers (grease-proof paper) and to the thin ebonite illustrated in figure 7, except that in the latter case the variation with electric stress was small so that for ebonite E should be deleted from equation (15.1).

---

[*] In the diagrams the product $\epsilon \tan \delta$ is plotted, where ϵ is the permittivity and δ the loss angle. This product is proportional to the conductivity σ since ϵ, δ σ

$$\sigma = f \epsilon \tan \delta \times 10^9 / 2C^2 \text{ mho/cm.},$$

where f is the frequency and C the velocity of light. f, C

Thin dry papers such as condenser tissue gave no appreciable effect of electric stress up to the breakdown range and only a slight variation with temperature. It will be mentioned later that with paper there is the possibility that the electric field may induce ionization in the pores of the material. This possibility does not probably extend to condenser tissue owing to its thinness*. A condenser tissue experimentally impregnated with a paraffin wax showed a variation similar to that denoted by equation (15.2).

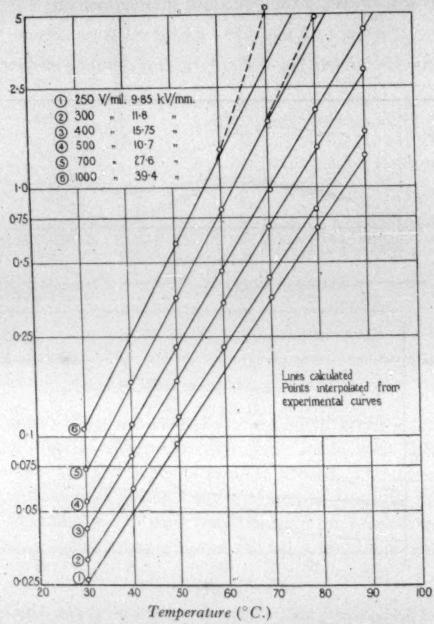


Figure 4. Effect of stress and temperature on specific loss of cellulose acetate, 0·193 mm. thick. (Conditioned 1 hour at 30° C.)

D.-c. conductivity. D.-c. conductivity decreases with time, and two hypotheses are commonly proposed in this connection: the polarization theory which considers the initial conductivity, before polarization has developed, to be the true value, and the absorption theory in which the final conductivity gives the normal conduction. Since from other investigations the polarization potential appears not to exceed a few volts and since the initial values of conductivity were found in the

---

* Hartshorn and Ward(5) found a sudden increase in the conductivity of condenser tissue at a field of 8·5 kV./mm. The difference may be due to the differences in the form of the electrodes or in the materials employed.

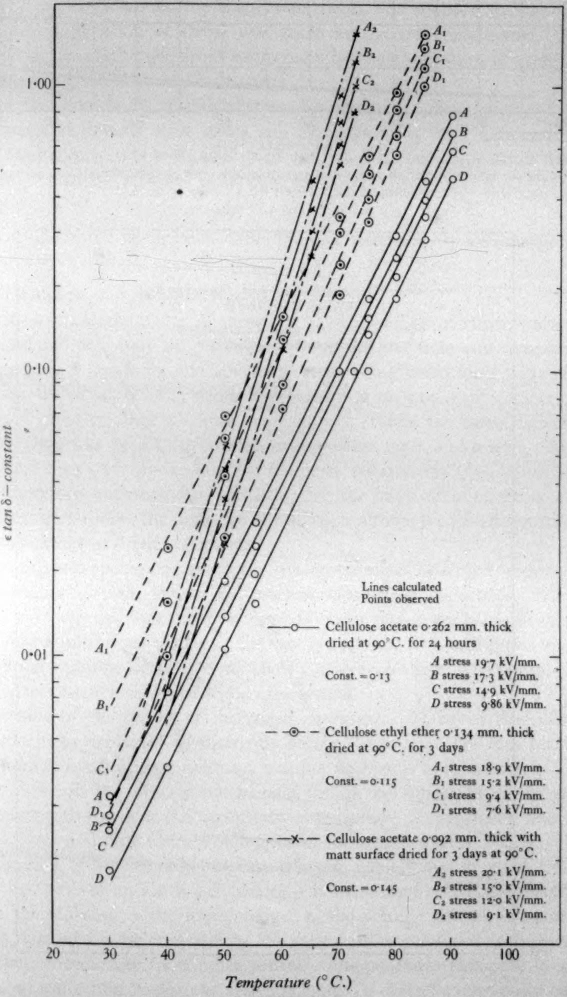


Figure 5. Energy-losses of cellulose compounds with slight traces of moisture.

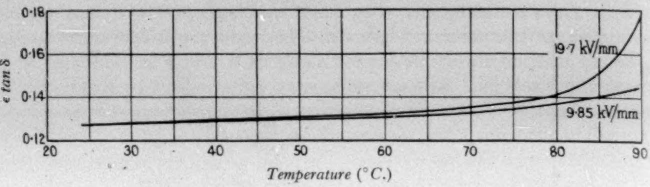


Figure 6. Effect of temperature and stress on the energy-loss of cellulose acetate, 0·265 mm. thick. (Dried for 14 days at 90° C.)

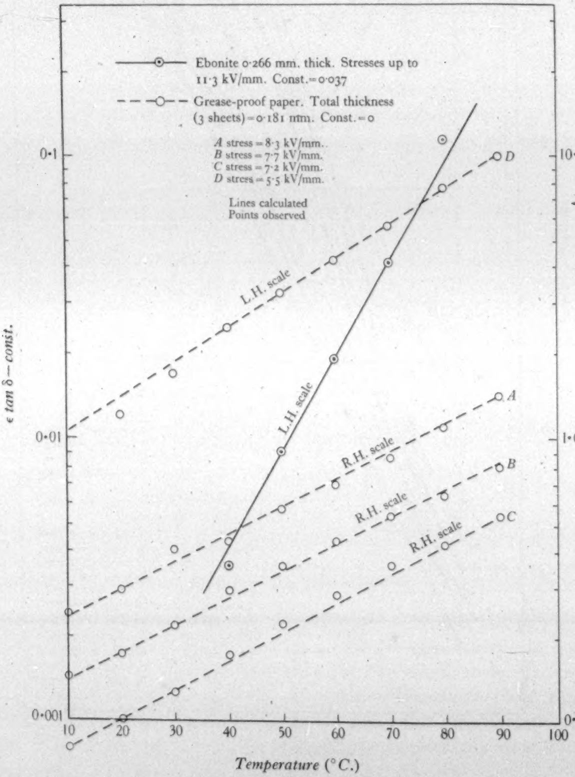


Figure 7. Energy-losses of special thin ebonite and grease-proof paper with slight traces of moisture. (Conditioned at 90° C.)

present tests to be variable and uncertain, while the final values could be repeated over a cycle of electric stress, the absorption theory was adopted and the value for 100 minutes' application of stress was taken as the true conduction. Previous electrification was discharged from a specimen by the application of an alternating field which was gradually reduced to zero.

These conditions being employed it was found that the conductivity of cellulose acetate increased linearly with stress and logarithmically with temperature. The effect of temperature was somewhat complicated. The best expression was found in one fairly extensive series of tests to be as follows:

$$\sigma = Ae^{\gamma_0 \theta} + (\theta - \theta_0)^{\gamma}\, aE \qquad \ldots\ldots(16).$$

This law is illustrated in figure 8. Thus the d.-c. conductivity also obeys the law

$$\sigma = \sigma_0 + aE \qquad \ldots\ldots(16a),$$

but σ_0, in addition to a, is a function of temperature.

As with a.-c., slight traces of moisture have a very great effect; variations of the order of 10^4 to 1 may be produced by varying the time and temperature of conditioning. For example, one specimen after being dried for 1 hour at 30° C. gave, when tested at 90° C., values for σ_0 and a of $5{\cdot}07 \times 10^{-11}$ mho/cm. and $0{\cdot}445 \times 10^{-11}$ mho/cm. per kV./mm. respectively. When the same dielectric was dried for 97 hours at 75° C. the corresponding values for σ_0 and a were $1{\cdot}84 \times 10^{-14}$ mho/cm. and $0{\cdot}14 \times 10^{-14}$ mho/cm. per kV./mm. respectively.

It is, therefore, reasonable to conclude that the large effect of stress and temperature is mainly due to the presence of moisture, although such effects still appear to a limited extent in dried specimens.

Unconditioned specimens show the hysteresis effect observed with ascending and descending electric stress for moisture-absorbing dielectrics. Metastable equilibria with the ambient medium seem to occur which are liable to move suddenly to new equilibrium values as the test is continued. Conditioned specimens are in internal equilibrium and do not show a hysteresis effect. The decay with time of the absorption current seems to be exponential.

Application of the theory of ionization coefficients. Many of the phenomena mentioned can be explained qualitatively on the theory that ions may form others by their motion in an intense field in a manner analogous to ionization by collision in gases, although the precise mechanism cannot be suggested at the moment. It is of interest to consider the quantitative application of this theory and some of the possibilities and difficulties which arise in connexion with it.

For the sake of brevity only one case is taken, namely that in which the ionization
per unit path of an ion is μE, where μ is a constant and E the electric field. μ, E
Let u be the mobility, ϵ the ionic charge, n_0 the volume density of spontaneous u, ϵ, n_0
ionization, assumed to be constant, n_0' the ionic density produced at the electrodes, n_0'
and n_0'' the ionic density from other sources. Suppose also that only ions of one n_0''
sign have an ionization coefficient. Then solving Townsend's continuity equations
for ionization by collision in gases, we have, V being the applied e.m.f., V

$$\sigma = (n_0\, u\epsilon/\mu E)\,(e^{\mu V} - 1) + n_0'\, u\epsilon^{\mu V} + n_0''\, u\epsilon \qquad \ldots\ldots(17).$$

If only the first term containing E is retained, this reduces to

$$\sigma = \sigma_0 + aE \qquad \text{......(18)},$$

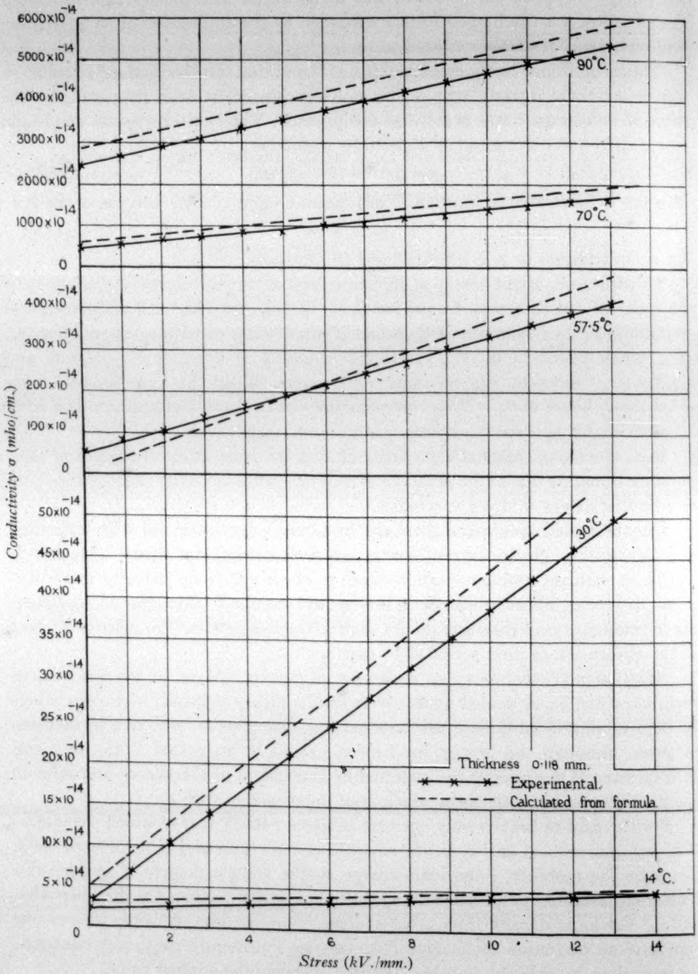


Figure 8. Relation between conductivity, stress and temperature for cellulose acetate containing moisture. (Conditioned 1 hour at 30° C.)

which has been encountered experimentally above. We observe that σ depends on the thickness, but with the materials under consideration other variations with thickness mask such phenomena. It is probable that the third term may be neglected in most dielectrics, particularly if they are moist.

With a.-c. the corresponding conductivity σ' is given by σ'

$$\sigma' = \frac{n_0 u\epsilon}{\mu E} \sqrt{2} . I_1 (\mu V \sqrt{2}) + \frac{8 n_0 u\epsilon \mu V}{\pi E} \sum_0^\infty \frac{(n+1\,!)^2 (2\mu V\sqrt{2})^{2n+1}}{(2n+2)!\,(2n+3)!}$$

$$+ n_0' u\epsilon \left[I_0 (\mu V\sqrt{2}) - \frac{1}{\mu V \sqrt{2}} I_1 (\mu V\sqrt{2}) + \frac{8}{\pi} \sum_0^\infty \frac{(n+1\,!)^2 (2\mu V\sqrt{2})^{2n+1}}{(2n+1)!\,(2n+3)!} \right] + n_0'' u\epsilon \quad \ldots\ldots(19).$$

E and V are r.-m.-s. values, I_0 and I_1 are Bessel functions of the first kind, with imaginary argument and of zero and first order respectively. The summation terms arise if the d.-c. relations are true for instantaneous currents and e.m.f.s independently of field-direction, i.e. for the symmetrical case. If, however, when the field is reversed, the d.-c. equations hold with E negative, then the summation terms are absent: this is the asymmetrical case. In either case, however, if only the first term is retained, we have I_0, I_1

$$\sigma' = \sigma_0' + a'E \quad \ldots\ldots(20),$$

which has been encountered experimentally above.

We note that experimentally the first term shows extremely wide variations depending on the moisture present. This is to be anticipated theoretically since this term depends on the initial ionization provided by electrolysis of the moisture. With a.-c. there is an additional term, due to absorption, which is fairly constant so that the variations are not very great and cannot be directly correlated with d.-c. values. It may be noted, however, that the difference is less at higher temperatures and higher moisture-contents where the absorption term is of less importance.

The calculation of μ is determined by the ratio a/σ_0 or a'/σ_0' and depends on the relative importance of n_0, n_0' and n_0''. The latter may usually be neglected, and in the cases considered the order of μ is affected in about the ratio of 2:1, according to whether n_0 and n_0' is the determining factor. Since it is only desired to give the order of μ, n_0' is neglected in order to give a basis of calculation, as also are the summation terms with a.-c.* If μ is constant then an ion gives on an average one

---

* It may be remarked that this type of conduction causes harmonics in the current wave. These have been observed by Gemant(6) when Ohm's law is departed from, but quantitative application cannot be made. If the field is taken as $E\sqrt{2} . \sin \omega t$, then the current i for the case considered is given by i

$$i = b_0 + \sum_1^\infty a_n \sin n\omega t + b_n \cos n\omega t.$$

For the symmetrical case

$$b_n = 0, \quad a_n = (-1)^{\frac{n+1}{2}} \left[I_n (x) + \sum_{k=1}^{k=\infty} (\tfrac{1}{2}x)^{2k} / \Gamma(2k+n+2)\, \Gamma(2k-n+2) \right].$$

For the asymmetrical case

$$4b_0 = x \frac{d}{dx} x^2 I_0 (x), \quad b_{2n} = (-1)^n\, 2x I'_{2n} (x), \quad a_n = (-1)^{n+1}\, 2x\, I'_{2n-1} (x),$$

where $x = \mu V \sqrt{2}$, i.e. x is the peak applied potential divided by the ionization potential. Philipoff(7) has given oscillograms showing the development of harmonics during thermal breakdown of the type ending in an explosive discharge.

additional pair of ions in falling through a constant potential-difference of $1/\mu$. This may be considered as a species of ionization potential and is so defined in table 1.

Table 1

Experimenter	Material	Type of current	Temperature (° C.)	Ionization potential (V.)
	Cellulose acetate:			
Present authors	0·0086 mm., dried at 30° C. 1 hour	d.-c.	51·5	260
		a.-c.	51·5	315
,,	0·0086 mm., dried at 30° C. 1 hour	d.-c.	90	480
		a.-c.	90	190
,,	0·0086 mm., dried at 75° C. 8 hours	d.-c.	90	310
,,	0·0086 mm., dried at 75° C. 73 hours	d.-c.	90	435
,,	0·0086 mm., dried at 75° C. 97 hours	d.-c.	90	550
,,	0·0118 mm., dried at 30° C. 1 hour	d.-c.	14	500
,,	0·0118 mm., dried at 30° C. 1 hour	d.-c.	30	98
,,	0·0118 mm., dried at 30° C. 1 hour	d.-c.	51·5	110
,,	0·0118 mm., dried at 30° C. 1 hour	d.-c.	70	220
,,	0·0118 mm., dried at 30° C. 1 hour	d.-c.	90	370
Hartshorn and Rushton	0·0118 mm., unconditioned	d.-c.	21	180
,,	0·0118 mm., dried at 90° C.	d.-c.	25	130
Present authors	0·0191 mm., dried at 40° C. 1 hour	d.-c.	90	95
,,	0·0191 mm., dried at 30° C. 1 hour	d.-c.	90	120
,,	0·0191 mm., dried at 20° C. 1 hour	d.-c.	90	110
,,	0·0191 mm., dried at 11° C. 1 hour	d.-c.	90	65
,,	0·0191 mm., dried at 30° C. 1 hour	a.-c.	90	210
	Varnished cloth:			
Hartshorn	L 105, dried at 90° C.	50 c./sec.	20	510
,,	L 105, dried at 90° C.	100 c./sec.	20	200
,,	L 105, dried at 90° C.	d.-c.	20	650
,,	L 101, dried at 90° C.	100 c./sec.	20	950
,,	L 101, dried at 90° C.	800 c./sec.	20	200
,,	L 101, dried at 90° C.	d.-c.	20	830

The results indicate that the ionization potential, though variable, does not change characteristically with moisture-content and with type of current, though there may be an increase with temperature. Furthermore the variation, though large on an absolute scale, is small compared with the very wide range of variation of the conductivity in the same conditions. It may be noted that the results obtained with varnished cloth by Hartshorn were fitted to a curve of the form

$$\sigma' = \sigma_0' + a'E^2 + b'E^4 \qquad \text{......(21)},$$

although the last term was not appreciable with cellulose acetate. This result is in favour of what we have called the asymmetrical case.

The linear variation with stress or a variation of the kind indicated by equation (21) has been found by other experimenters such as Poole and Schiller and also in radio-frequency experiments carried out for the E.R.A. In many cases, however, an alternative form

$$\sigma' = \sigma_0' + aE^{\gamma} \qquad \text{......(22)},$$

proposed by Hartshorn, also agrees with experimental results.

The variation with temperature might be expected to follow the law $e^{-\gamma/T}$ rather than $e^{\gamma\theta}$, where T is the absolute temperature.

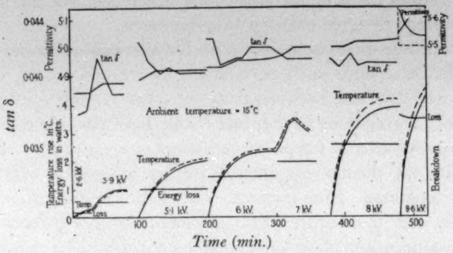
Over small ranges there is not very much difference, but the latter form was found on the average to agree better with experiment.

One may conclude that there is a case for the application of the theory of ionization coefficients, since such correlations as are effected by it between the results for a.-c. and d.-c. and between those for other variable conditions represent an advance. Such correlations have hitherto only been possible to some extent for absorption energy-losses and for poor conductors obeying Ohm's law. It is clear, nevertheless, that the phenomena require further analysis in order that various factors may be separated. For instance, with a.-c. the absorption loss must be subtracted while with d.-c. other forms of conduction are present besides those envisaged. The conception of an ionization potential will be considered later in connection with electric strength.

§ 3. PHENOMENA OF INSTABILITY

Breakdown processes. Many materials, particularly if unconditioned, become chemically unstable under prolonged high electric stress, so that although thermal instability may occur the phenomena cannot be predicted from a theory based on stable properties, since the electrical properties do not persist. The materials for which formulae were developed for the stable electrical properties in § 2 were, however, stable in so far that, when they were exposed to prolonged electric stress in the apparatus of figure 2, their electrical properties agreed with the formulae already given, provided the temperature of the specimen, as measured by the thermocouples, had approached fairly close to the final value corresponding to the applied stress. This agreement was also exhibited, to a somewhat lower degree of accuracy, for all electric stresses such that the rate of temperature-rise decreased to a point at which it did not vary greatly during a given measurement, even though subsequently the specimen became unstable under the same electric stress. Further, the energy-losses determined from the temperature-rise when thermally stable, agreed with the electrical measurement of a.-c. conductivity within the accuracy of the thermal determinations which, owing to the low temperature-rises, was about 10 per cent. Thus it may be assumed that substantially all the electrical energy is converted into heat.

Three types of behaviour at breakdown can be distinguished. (i) The temperature-rise, if appreciable, is either stable or increasing slowly at breakdown, which takes the form of an explosive discharge. The conductivity is also stable or else increasing slowly in dependence on the temperature. This type of behaviour was shown by all materials at low temperatures and by dried materials at all the temperatures employed. With ebonite and paper the temperature-rise was very small owing to the very low conductivity. Figure 9 illustrates the behaviour of cellulose acetate containing moisture at an ambient temperature of 15° C. For dry ebonite, paper, and cellulose compounds only a slight temperature-rise and a slight variation of conductivity occurred right up to breakdown. (ii) The conductivity and temperature become unstable beyond a certain electric stress, but an explosive discharge occurs while the conductivity and temperature are still of normal magnitude. This is illustrated in figure 10 for cellulose ethyl ether. (iii) The conductivity and temperature become unstable and the former rises to values such that insulating properties are lost. The temperature may reach the limit of chemical stability or a discharge may occur. An exaggerated case is shown in figure 11 which occurred at a high ambient temperature.



Mean temperature rise of l.-v. plates ———. Temperature rise of h.-v. plate -----.

Figure 9. Sustained test. Cellulose acetate, 0·26 mm. thick, containing moisture.

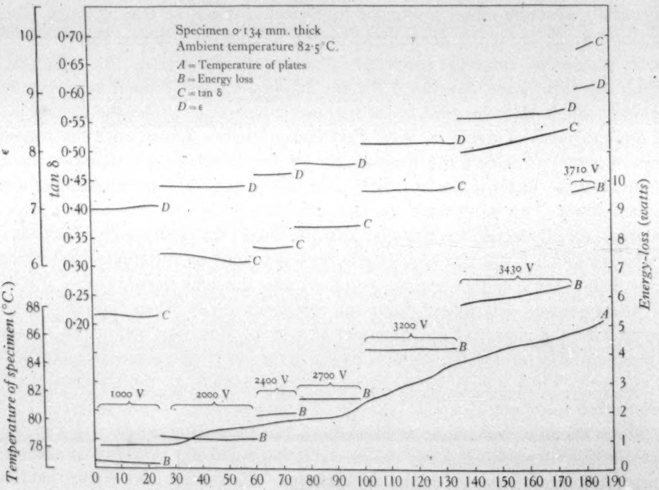


Figure 10. Sustained test on cellulose ethyl ether containing slight traces of moisture.

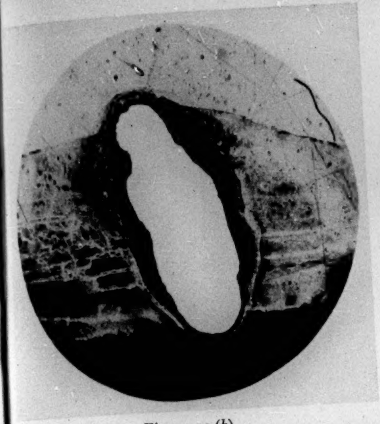


Figure 12 (*b*).

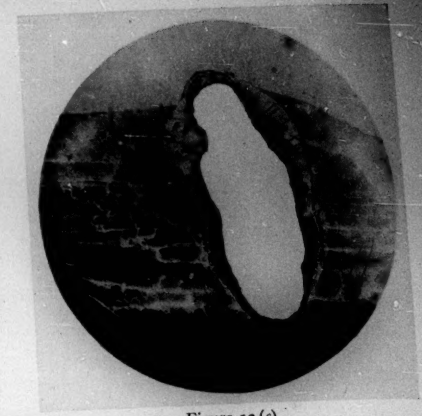


Figure 12 (*c*).

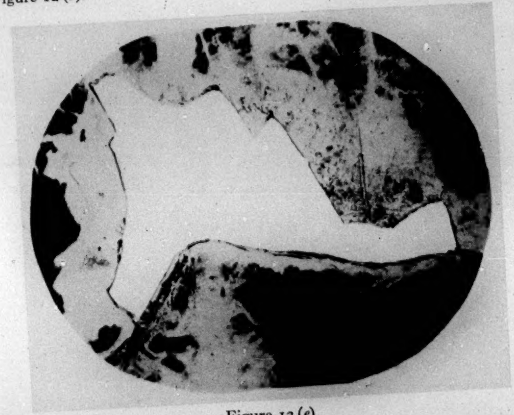


Figure 12 (*e*).

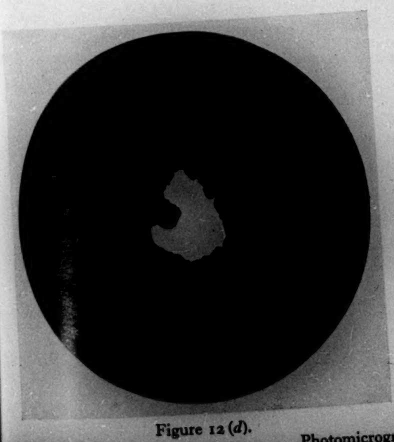


Figure 12 (*d*).

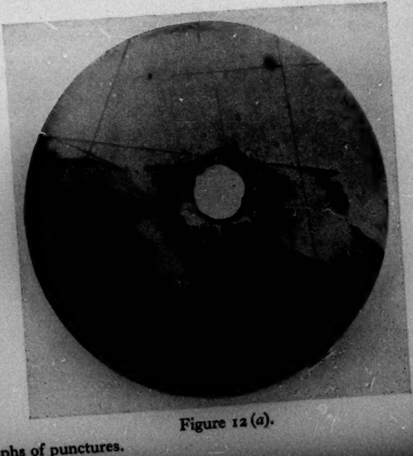


Figure 12 (*a*).

Photomicrographs of punctures.

I

air

inte

the

spr

cor

test

as

the

Figure 12 shows photomicrographs of punctures. (*b*) and (*c*) were taken with air and oil immersion respectively, and indicate that the blackening is due to internal reflection and not charring. (*a*), (*b*) and (*c*) correspond to type (i) above, the discharge in (*a*) being of very short duration. In (*b*) the initial circular puncture spread because the discharge was allowed to continue for a short time. (*d*) and (*e*) correspond to types (ii) and (iii). Similar results were obtained with short-time tests of electric strength; punctures similar to figure 12(*a*) or (*d*) occurred according as thermal instability did not or did take place. The shorter the time of discharge the smaller does the non-thermal puncture become, down to the point where it is scarcely visible. Type (*a*) was often situated at the edge of the electrode, but type (*d*) was nearly always inside. A. M. Thomas[(8)] has shown that punctures of type (*a*) appear to occur when extraneous discharges are eliminated. In the present instance excessive precautions such as were used by him were not employed. Nevertheless, the appearance is similar.

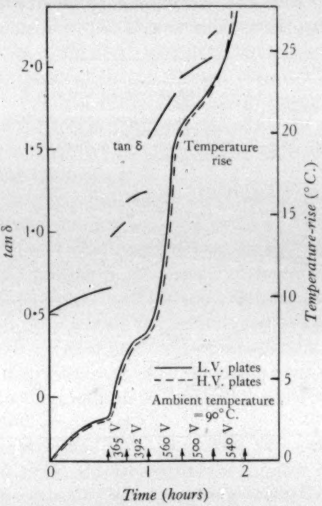


Figure 11. Instability of cellulose acetate 0·05 mm. thick, containing moisture. (Conditioned 1 hour at 30° C.)

Variation and limits of thermal electric strength. The theory given in § 1 shows that for sustained tests the thermal electric strength should be calculable and should follow approximately a logarithmic law with respect to ambient temperature, while short-time tests should give curves parallel, on logarithmic paper, to the curve for sustained tests, the height of the curve increasing with decreasing time of stress. Figure 13 illustrates this from experiments on partially dried cellulose acetate. The curves mainly consist of two portions, one of which shows little variation with temperature and refers to lower temperatures, shorter times, and drier and thicker specimens, while the second portion refers to higher temperatures, longer times, and moister and thinner specimens. The second portion obeys the principles of thermal instability set out above and gives quantitative agreement for the sustained tests. The first portion we can relate to a non-thermal type of discharge. Thus the laws of thermal instability are obeyed provided the thermal electric strength is less than the non-thermal electric strength for the conditions of the test. The non-thermal electric strength will be considered later.

The thermal range is associated with instability phenomana of types (ii) and (iii) described above, and with the corresponding micrographical appearance of punctures. The latter applies to short-time as well as sustained tests. The effect of decreasing the time of stress is to move the transition point to higher temperatures. The same result applies to decrease of moisture-content as is shown in figure 13 for tests on films 0·19 mm. thick. Thermal instability did not come within the range of the experiments with the completely dried specimen. Owing to variations in method of manufacture for different thicknesses, the effect of thickness with cellulose acetate films does not follow a simple law. A similar complexity was found with varnished cloth in consequence of increase of energy-loss with the number of layers. With cellulose acetate the thermal electric strength should be inversely proportional to the cube root of the thickness. Actually, owing to the greater energy-loss in thin specimens, the thermal electric strength increases with thickness. With varnished cloth the thermal electric strength should normally be inversely proportional to the square root of the thickness, but actually it decreased more rapidly than this owing to an increase of energy-loss with number of layers.

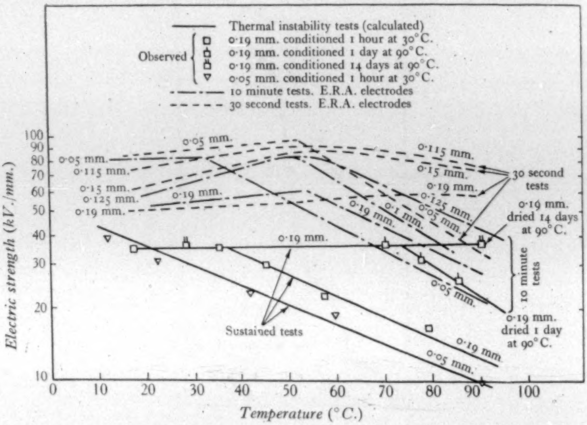


Figure 13. Electric strength of cellulose acetate. (Conditioned 1 hour at 30° C. except where the contrary is stated.)

We have neglected the maximum and minimum of energy-loss which occurs below 30° C.—the so-called *V* curve. It has no influence on the present results and thermal breakdown can only be realized at higher temperatures. A detailed discussion has been given elsewhere and it need only be remarked that a slightly anomalous electric-strength curve is produced, the anomalous curve giving apparently lower values of electric strength than the normal curve.

Cellulose acetate with a matt surface and cellulose ethyl ether conditioned at 90° C. gave results similar to those given by cellulose acetate, the transition temperatures being different owing to differences in the energy-losses and to a smaller degree in the non-thermal electric strength. Thermal breakdown occurred above 63° C., in agreement with theory for the matted material with sustained tests. For cellulose ethyl ether, the theoretical transition temperature was 54° C. on the assumption that the non-thermal electric strength was 39 kV./mm. The specimens were not, however, very uniform in this respect and in some sustained tests a non-thermal electric strength of 29 kV./mm. was indicated. In such circumstances the theoretical transition temperature was 84·5° C. The majority of the tests, however, confirmed the theoretical value of 54° C.

With dried ebonite the theoretical transition temperatures in sustained tests fell outside the range of the experiments, being 94, 97·5 and 101° C. with thicknesses of 0·52, 0·385 and 0·275 mm. respectively. Owing to the low non-thermal electric strength of dried grease-proof paper, the transition temperature is very high, being 206° C. for 3 sheets and 143·5° C. for 15 sheets. Condenser tissue paper when dry is very slightly affected by stress or temperature up to the non-thermal electric strength, and accordingly thermal breakdown cannot occur except with very great thicknesses where the thermal resistance of the paper itself is much greater. In all these cases it was verified that breakdown through thermal instability did not occur at the lower experimental temperatures.

Non-thermal electric strength. It is of interest to consider the nature of the electric strength which limits the application of thermal laws, although this investigation did not form part of the original programme and is being pursued by other means. There is an initial difficulty in determining the influence of edge effects and extraneous discharges. According to A. M. Thomas it is necessary to utilize a penetrating ambient medium of high electric strength, such as xylene under pressure, to eliminate these effects, and in such circumstances the electric strength of, for example, cellulose-acetate film is three or more times the values here observed. On the other hand the type of puncture resembles the type which Thomas considers characteristic of the elimination of edge effects. Figure 14 shows the effect of thickness for non-thermal and thermal breakdown with cellulose acetate. The former follows the parabolic law, the thickness being inversely proportional to the square of the average electric strength; this is the law arising from a constant limiting maximum field-strength applied to the field between two thin discs. The effect of introducing a graphite contact film is to increase the electric strength by about 10 per cent. In the sustained tests breakdown occurred more often within the guard ring than at the edge, following a similar law. With ebonite, breakdowns fell into two groups according to whether the puncture was at the edge of the electrode or within it. In both groups the electric strength was independent of thickness. Tests on different numbers of layers of paper, both tissue and grease-proof, gave a constant electric strength. It is not clear, therefore, how far these values are characteristic of the material or whether the field-distribution of ambient medium exert an important effect, and investigations are proceeding with respect to these questions.

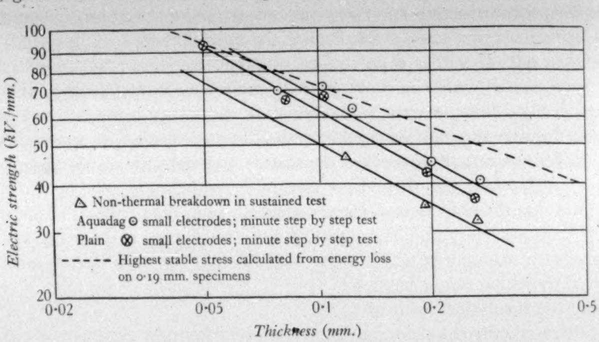


Figure 14. Electric strength of cellulose-acetate films of various thicknesses at 20° C.

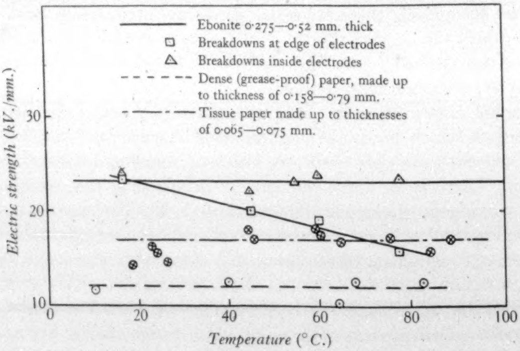


Figure 15. Non-thermal breakdown in sustained tests.

From figure 13 it is seen that temperature has not a very great effect on the non-thermal electric strength of cellulose-acetate films. A slight increase with temperature appears possible, but usually it is not great in comparison with the inherent variability of this quantity. Figure 15 shows that for ebonite the electric strength for internal breakdowns is constant, while for edge breakdowns it decreases with temperature. The results of sustained tests on grease-proof paper and condenser tissue paper were also unaffected by temperature, but short-time tests gave somewhat different results. The electric strength of condenser tissue was constant, while the electric strength of grease-proof paper decreased slightly with temperature whether tested in air, figure 16, or in oil, figure 17, although impregnation with oil increases the electric strength considerably to approximately the same value as when the tissue paper is impregnated with paraffin wax. The electric strength of a layer of air of the same thickness as a single sheet of tissue paper is of the order of

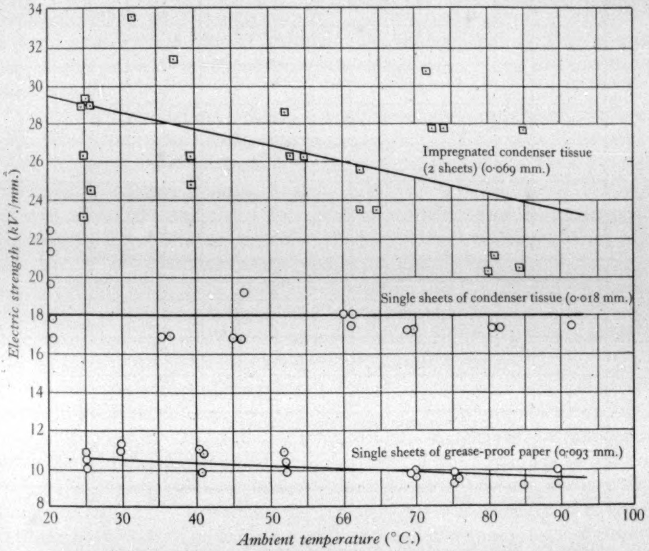


Figure 16. Short-time electric strength of paper in air. (Total time 1 minute.)

30 to 40 kV./mm., while for a single sheet of grease-proof paper, it would be about 10 kV./mm. Thus there seems reason to believe that the linear decrease with temperature is probably associated with the intervention of the ambient medium at edges or at interstices in the structure. The effect is difficult to separate from the inherent variations.

The non-thermal electric strength is not much affected by small quantities of moisture, as for example in the experiments on cellulose acetate films. The effect of duration of test was not studied in detail except that it was noted, as is well-known, that the electric strength decreases with duration of test. This effect is, for durations greater than 10 seconds, much less than the corresponding effect on the thermal electric strength. This finding is in agreement with more recent work on dielectrics such as that of Inge and Walther, and it may be concluded that the non-thermal electric strength of dielectrics does not normally vary greatly with duration of test in the absence of strong external ionization in the ambient medium.

It is also to be noted that corrugation of the surface of cellulose acetate, as in matt films, had no appreciable effect on the electric strength. This is to be expected since the electric strength of fine channels when filled with air or oil would be greater than that of the solid material.

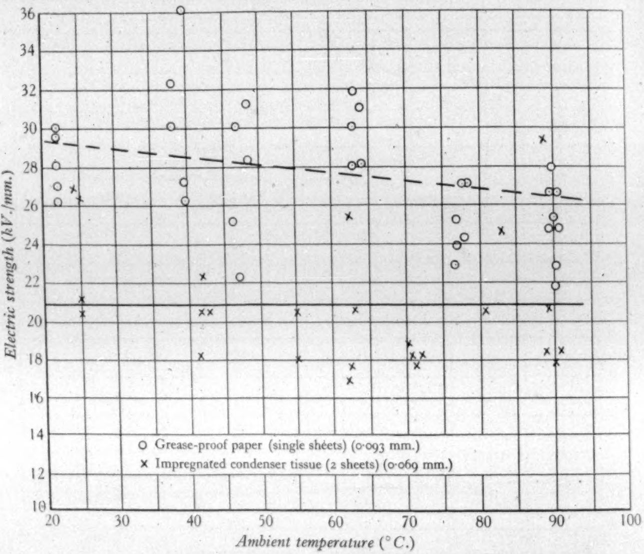


Figure 17. Short-time electric strength of paper immersed in transformer oil. (Total time 1 minute.)

Tests made on paper with continuous voltages gave the same results as those made under the same conditions with alternating voltages having peak values equal to the values of the continuous voltages.

Modern views of intrinsic electric strength. Although the present paper deals only indirectly with the intrinsic electric strength, which enters as an upper limit, it is of interest to consider recent theories dealing with this subject since it appears likely that two or more groups of effects are involved.

Joffe(9) arrives at ionization potentials of the order of 100 or 200 volts, although he finds a decrease with field-strength not observed by the present authors. Such values, and those given in this paper, fit in with his theory of transition from ionic breakdown (about 10^6 V./cm.) to structural breakdown (10^8 V./cm.) from thicknesses of 10^{-3} cm. to those of 10^{-2} cm., since the voltage across the specimen has reached the minimum for ionization and cannot fall further until another type of failure occurs. This, however, is an over-simplification. Ionization can occur at small thicknesses at lower voltages than are possible with these values of ionization potential: ionization by ions of one sign does not become unstable without the intervention of some other factor, such as ionization by ions of opposite sign: ionic energies calculated from structure are very much less than these ionization potentials; and finally some doubt attached to Joffe's electric-strength values of 10^8 V./cm. Examination of actual discharge currents also supports the view that the high ionization potentials are not physically real, but are statistical and average effects due to energy-dissipation without ionization; and although an ionic form of breakdown probably exists, it is related to current-density phenomena and not directly to the structure.

Calculations of ionization potentials of polar molecules usually give low values⁽⁹⁾. Wideroe⁽¹⁰⁾ arrived at ionic energies from the heats of sublimation and dissociation for rocksalt, giving 2·8 electron-volts for the sodium ion and 11·6 for the chlorine ion. Hippel⁽¹¹⁾ indicates a proportionality between the voltage-drop in a lattice element at breakdown and the energy-difference in electron-volts corresponding to the Reststrahlen of different crystals. His view is that an electron raised to the highest level of an incomplete energy band by the electric field may, by falling to a lower level in the band, yield energy to an electron in a lower complete band, so that two free electrons are produced. The ionization potentials so deduced from the voltage-drop per unit element would be only of the order of a fraction of a volt, the figures for electric strength being between 10^5 and 10^6 V./cm.

A calculation has recently been made by Zener⁽¹²⁾ of the rate of escape of electrons from a complete band through an energy-difference into an incomplete band under the action of an electric field. He finds that this rate of escape is so rapidly accelerated as the field increases beyond, say, the order of 10^6 V./cm., as to constitute breakdown, assuming a lattice constant of the order of 3×10^{-8} cm. and an energy-difference of, say, 2 V. If we take the equations for semi-conduction in crystals worked out by Wilson⁽¹³⁾ on the same basis of closed and open groups of electrons in solid lattices and then consider that a species of Stark effect occurs giving new intermediate energy-levels and thus inducing metallic conduction, we find that if the same order of constant occurs as in gases, conduction in a field of the order of 10^6 V./cm. should occur if the energy-level displacement is to be of the order of 0·01 V., which is about the order of energy-differences exhibited with some semi-conductors which are insulators at low temperatures. This view however leads to an interaction of temperature and field-strength giving a term of the form exp $(\lambda E/kT)$ in the conductivity. Thus the lower the temperature the greater the relative effect of stress, so that although an exponential variation of conductivity with stress is very frequent, it can hardly be ascribed to such a mechanism since the corresponding temperature effect has not been observed.

At present the general conclusion appears to be that the final limiting electric strength is probably nearer 10^6 or 10^7 V./cm. than the 10^8 V./cm. given by Joffe, and is associated with energy-differences possibly as low as hundredths of a volt but probably not greater than a few volts. This structural breakdown should be sudden in character giving a phase of electronic conduction, at any rate initially, which is unlikely to be related to the conduction preceding breakdown or to be greatly affected by time or temperature over a wide range. An ionic form of breakdown, comprising a kind of electrolytic decomposition with induced ionization, may occur at lower field-strengths and will be related to the conduction immediately preceding breakdown. The ionic type of electric strength should show time effects at longer times than the structural type but should not necessarily be greatly affected by temperature.

In the present tests the times were too long to enable such distinctions to be made experimentally; the breakdown might have been of either type. In addition it is not yet completely known how to avoid external effects, imperfections in the material, or the effect of a mosaic structure if present. On the other hand the fact that electric strengths of 10^6 V./cm. and above can sometimes be realized with suitable precautions with many materials indicates that these difficulties may not be so great in the electrical problem as in the corresponding mechanical problem. Work is proceeding on the non-thermal type of breakdown at shorter times to attempt a further separation of breakdown types.

§ 4. CONCLUSIONS

(i) Provided that a sufficiently wide range of external conditions is available a dielectric will exhibit at least two types of breakdown—a thermal and a non-thermal type. In the case of thermal breakdown the electric strength corresponding to a long time of stress can be calculated with adequate accuracy from the conductivity under stable conditions, when the external thermal conditions and also the thermal properties of the dielectric are known. Provided the electric strength so computed is less than the non-thermal electric strength, the calculation will be valid whether breakdown takes the form of a progressive loss of insulating properties or the form of a sudden explosive discharge. The thermal electric strength depends on the electrical processes preceding breakdown only in so far as these processes contribute to the generation of heat in the dielectric. Joint electrical and thermal effects of the type suggested by Rogowski and others were definitely absent in the present work.

(ii) If a stronger field is applied so that breakdown takes place in a short time the theory of thermal instability may be extended to give the electric stress corresponding to this time of application. The quantitative agreement between theory and experiment is not then so accurate, but if the variation of electric strength with time is known for a given temperature the variation at other temperatures can be theoretically deduced therefrom with sufficient accuracy.

(iii) A transition temperature exists for any given time of application of electric stress and given external conditions. Below this temperature non-thermal breakdown occurs, above it the electric strength obeys the theory of thermal instability. The range of application of the latter theory is accordingly confined largely to higher temperatures, thicker specimens, poorer dielectrics, longer times of application, or electrode arrangements of higher thermal resistance.

(iv) Although there is evidence from the phenomenon of a.-c. and d.-c. conductivity to the effect that ionic effects occur, and although ionization potentials which give a fairly simple account of the phenomena can be deduced, yet it is doubtful whether these stable processes bear a direct relation to any of the types of non-thermal breakdown. The experimental methods employed showed no evidence of a factor which would be capable of producing ionic instability and also determine the conductivity at lower field strengths.

(v) The ionization potentials deduced are much greater than those which can be obtained from considerations of solid structure, and they have not the same physical significance as the corresponding quantities in gases. On the other hand the electronic theory of intrinsic breakdown does not appear to be capable of explaining the stable phenomena.

(vi) The electronic form of non-thermal breakdown should be independent of temperature, thickness of dielectric, time of application of stress, and humidity. The same appears to be true of the ionic form, except that this form is affected to some extent by the time of application of stress.

§ 5. ACKNOWLEDGMENT

The authors wish to thank Prof. MacGregor-Morris, in whose laboratory the experimental work was carried out, for the facilities provided and for his help during the investigation.

REFERENCES

(1) *Report of the British Electrical and Allied Industries Research Association*, Ref. L/T 69.

(2) WHITEHEAD. *Dielectric Phenomena*, **3** (Benn).

(3) FOCK. *Arch. Elektrotech.* **19**, 71 (1927).

(4) WAGNER. *J. Amer. Inst. elect. Engrs*, **41**, 283; *S.B. Akad. Wiss. Wien*, **29**, 438 (1922).

(5) HARTSHORN and WARD. *Report of the British Electrical and Allied Industries Research Association*, Ref. L/T 48.

(6) GEMANT. *Arch. Elektrotech.* **23**, pt. 6, p. 683.

(7) PHILIPOFF. *Z. tech. Phys.* **14**, No. 1, p. 21 (1933).

(8) A. M. THOMAS. *Report of the British Electrical and Allied Industries Research Association*, Ref. L/T 57.

(9) JOFFE, see (2).

(10) WIDEROE. *Arch. Elektrotech.* **26**, pt. 9, p. 626 (1932).

(11) HIPPEL. *Z. Phys.* **67**, 707; **68**, 309; **75**, 145.

(12) ZENER. *Proc. roy. Soc.* A, **145**, 523 (1934).

(13) WILSON. *Proc. roy. Soc.* A, **133**, 458 (1931); **134**, 277 (1931).

(14) FOWLER. *Proc. roy. Soc.* A, **140**, 505 (1933).

DEMONSTRATION

THE USE OF A LIQUID SURFACE CARRYING RIPPLES AS A DIFFRACTION GRATING.

Given by R. C. Brown, Ph.D., B.Sc. Lecturer in Physics, University of London, University College.

Demonstration given on June 7, 1935.

It is well known that from measurements of the wave-lengths of ripples of known frequency the surface tension of the liquid upon which the waves are travelling can be calculated. The most usual method of arriving at the wave-length is that of direct measurement, either of progressive waves viewed stroboscopically or of stationary waves.

Kalähne⁽¹⁾ showed that it was possible to use a rippling surface as a diffraction grating and by measuring the angular separation of the spectra he was able to calculate the wave-length of his ripples, i.e. the grating-space. Stationary ripples were used but, owing to the enormous disparity between the velocity of the ripples and that of light, progressive waves will provide an equally good grating.

The following is a description of a very simple experiment designed to demonstrate the spectra produced. In figure 1, *C* and *T* represent the collimator and telescope respectively of a small spectrometer. They are adjusted for parallel light and arranged so that an image of the slit is seen in the telescope by reflection at almost grazing incidence from the surface of water contained in a half-plate developing dish. *D* represents the dipper consisting of a small glass plate attached to the diaphragm of a loud-speaker unit so that when the loud-speaker is actuated by a small valve oscillator the dipper vibrates vertically while keeping parallel to its own plane.

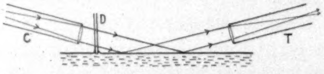


Figure 1.

The position of the dipper should be such that no reflection of light takes place from portions of the water-surface which are near enough to it to be curved as a result of capillarity. In order to effect this adjustment, it is convenient to restrict the light leaving the collimator by means of a small piece of card.

If the bench on which the experiment is being performed is liable to vibration, the dish may be held in a sling of rubber tubing in the way in which microphones are suspended. It may be necessary also to place a thickness of sponge rubber between the bench and the stand which carries the loud-speaker unit and dipper, for irregular disturbances originate from the dipper if this is not done.

ERRATA

"An investigation of the wet-and-dry-bulb hygrometer at low temperatures", by J. H. Awbery, B.A., B.Sc., F.Inst.P., and Ezer Griffiths, D.Sc., F.Inst.P., F.R.S., *Proc. Phys. Soc.* **47**, 699 (1935).

For table 5 substitute the following

Wet-bulb depression (° C.)	Dry-bulb temperature (° C.)						
		0	−4	−8	−12	−16	−18
0	Meteorological Office	100	96	92	89	—	—
	Marvin	100	96	92	89	86	84
	Assmann	100	97	93	89	86	84
	Present paper	100	97	89	75	66	62
1·0	Meteorological Office	81	73	63	53	38	—
	Marvin	81	74	65	54	41	32
	Assmann	81	74	65	55	41	32
	Present paper	80	74	62	49	37	33
2·0	Meteorological Office	61	50	35	15		
	Marvin	63	52	38	21		
	Assmann	63	52	38	21		
	Present paper	62	53	36	24		
3·0	Meteorological Office	43	27	—			
	Marvin	45	31	12			
	Assmann	46	31	—			
	Present paper	44	32	16			
4·0	Meteorological Office	25	—				
	Marvin	28	11				
	Assmann	28	—				
	Present paper	31	17				

a small valve oscillator the dipper vibrates vertically while keeping parallel to its own plane.

The position of the dipper should be such that no reflection of light takes place from portions of the water-surface which are near enough to it to be curved as a result of capillarity. In order to effect this adjustment, it is convenient to restrict the light leaving the collimator by means of a small piece of card.

If the bench on which the experiment is being performed is liable to vibration, the dish may be held in a sling of rubber tubing in the way in which microphones are suspended. It may be necessary also to place a thickness of sponge rubber between the bench and the stand which carries the loud-speaker unit and dipper, for irregular disturbances originate from the dipper if this is not done.

The production of standing waves by reflection from the sides of the dish gives rise to a continuous background of illumination and detracts from the clarity of the spectra. This can be avoided by using small amplitudes.

Good diffracted spectra are seen in the telescope when white light is used and when the frequency of the ripples is between 50 and 500 c./sec. Above this range it is difficult to produce waves of large enough amplitude to give any observable effects.

So far the author has not adopted this somewhat indirect method of measuring ripple wave-lengths, since it would not be superior to the direct method of measurement except at high frequencies, i.e., with wave-lengths too short to be measured accurately on a divided scale. For frequencies below 500 c./sec. wave-lengths are large enough to be measured directly with sufficient accuracy.

REFERENCE

(1) KALÄHNE. *Ann. der Phys.* **7**, 440 (1902).

REVIEWS OF BOOKS

The Principles of Quantum Mechanics, by P. A. M. DIRAC, Fellow of St John's College, Cambridge. Second edition. Pp. xi + 300. (Humphrey Milford: The Oxford University Press, 1935.) 17*s.* 6*d.* net.

The first edition of this book, which was published in 1930, marked a new epoch in the development of the quantum theory. In this second edition considerable changes have been made in the method of presentation, and a certain amount of additional matter has been included.

The earlier chapters, which built up the general theory of the representation of states by vectors and of observables by linear operators, have been almost completely rewritten and the method of presentation has been somewhat simplified. The theory nevertheless remains highly abstract in character and makes no concession to the physicist who wishes for an objective and realist interpretation of the fundamental axioms of this subject.

In the chapter on the equations of motion Dirac has incorporated some of his recent work on the form taken by the principle of least action in quantum theory, together with a section, which is all too brief, on the theory of the statistical ensemble.

The chapter on the relativistic theory of the electron includes Schrödinger's exact solution of the equations of motion of a free electron, together with the theory of the positron, the existence of which, it will be remembered, was predicted by Dirac over a year before it was discovered experimentally.

The last chapter in the book is entirely new and is devoted to the fascinating subject of field theory, due to Jordan and Pauli, Heisenberg and Born. The interest of this particular development is that it is concerned with the principal problem which so far has defied solution by the methods of the quantum theory. Indeed it appears that some essentially new principle is required to deal with the difficulties of micro-microphysics on the scale of electronic and nuclear dimensions.

The book is beautifully printed and produced by the Oxford Press. The reviewer has discovered only one misprint, namely on page 74, line 4, where an accent is missing from the δ symbol.

G. T.

The Measurement of Inductance, Capacitance, and Frequency, by ALBERT CAMPBELL and ERNEST C. CHILDS. Pp. xxiv + 488. (London: Macmillan and Co.) 30*s.* net.

It is a fortunate thing that Mr Albert Campbell has been able to tear himself away from his laboratory for a sufficient time to enable him, with the assistance of Mr Childs, to write this book. Probably no man living has had greater experience in the matters of which it treats than he, and in this work he succeeds in handing on some of the fruits of this experience to other workers.

The book may perhaps be said to be somewhat unbalanced inasmuch as the more modern work at radio frequencies is not presented in quite the same comprehensive and masterly manner as, say, vibration galvanometers, the construction, calculation and measurement of inductors, and condensers, the alternating-current potentiometer, and frequency-meters. But it must be a difficult matter to decide on the exact scope of such a work and incidentally to give it an adequate title. It gives much valuable information, not easily accessible elsewhere, on general electrical measurements. Almost the whole field of alternating-current measurements may be said to have been covered, including measurements at low, audio and radio frequencies, but there is no doubt that it will be found most useful by the man who is concerned mainly with audio-frequency measurements. Many interesting details concerning the history of the subject are touched upon. For example, many physicists must have wondered who coined the terms inductometer and variometer, and what is meant by them. Mr Campbell is a safe guide on such matters. The book should be noted by all physicists who are concerned with measurements involving the use of alternating currents, including work on dielectrics, magnetic materials, and electrolytes, as well as the more obvious measurements on condensers, inductances, and laboratory standards generally. The book is very clearly and concisely written, conveniently arranged and indexed, and in every way well adapted to its purpose. L. H.

Intermediate Physics, by C. J. SMITH, Ph.D., M.Sc., A.R.C.S. Second edition. Pp. xii + 900 with illustrations and diagrams. (London: E. Arnold and Co.) 16*s*. Also published in parts.

The original issue of this book was criticized for a certain lack of balance, but this has been corrected in this new edition. The section on electricity and magnetism has been thoroughly revised and amplified, and the opportunity has been taken of making small improvements and additions throughout.

The text now provides a very suitable background for the usual type of intermediate course, especially for the abler students. It will be particularly useful for those preparing for more advanced work to come, or aiming at scholarships. Ostensibly no previous knowledge of the subject is assumed, but the book does not appear to be suitable for an absolute beginner. It is mainly experimental in outlook, but does not profess to provide the detail which is looked for in a practical text-book. It pays little attention to the historical aspect of the subject—which is a pity, since this is too often under-valued.

Those familiar with the first edition will be interested to learn that the book can now be obtained in separate parts, five in number, the dearest being that on electricity and magnetism at 6*s*. Short mathematical tables are included, but numerical data of a physical nature are still inadequately provided, and the student must glean what information on magnitude he can from chance numbers in the problems. The theory of dimensions occupies a more prominent place than in the previous edition, and there is more scope for the student with good mathematical equipment. The diagrams are even better than before.

The general treatment is very sound, and pleasurably free from fallacies and doubtful "simplifications". This does not mean that there are no errors at all. Indeed, the author's ideas on laws, physical and penal, and on the physical meaning of negative magnification, for example, need revision; and here and there an example or a phrase is used before the way has been prepared for it in the text. But the discriminating teacher will easily remedy these defects, and will probably find the book no worse in this respect than most others of its size and scope. It is certainly very good value for money, and can be strongly recommended. A. J. W.

Through the Weather House, by R. A. WATSON-WATT. Pp. xi + 192. (London: Peter Davies, Ltd.). 7*s*. 6*d*. net.

Probably many Fellows of the Physical Society listened with no less pleasure than the general public to the popular series of talks on meteorology broadcast by Mr Watson Watt early in 1934. These lectures are now published in revised and expanded form under the title *Through the Weather House*, recalling the picturesque imagery with which Mr Watt adorned his fascinating tale of our atmosphere, its beauties, its wonders, and its problems. The book deserves and will assuredly gain a wide sale. S. C.

Fine Structure in Line Spectra and Nuclear Spin, by S. TOLANSKY. Pp. viii + 112. (Methuen's Monographs on Physical Subjects. London: Methuen & Co., 1935.) 3*s*. net.

In recent years optical spectroscopy has added much to our knowledge of the atomic nucleus. The magnitudes of the angular momenta and magnetic moments of a large number of nuclei have been derived from studies of hyperfine structure of atomic lines and of alternating intensities in the rotational structure of bands of homonuclear diatomic molecules, whilst for the recognition of stable isotopes and determinations of their relative masses and abundances the isotope effect in diatomic band spectra, especially, has supplemented the mass spectrograph method to a large and important extent. One of these fruitful branches of modern spectroscopy is lucidly described in this monograph, largely with the aid of vector and energy-level diagrams and reference to actual examples. An account of electron-spin structure leads up to that of nuclear-spin structure. Derived values of the nuclear-spin quantum number I and magnetic-moment factor $g(I)$ are tabulated and discussed, and theories of nuclear spin are outlined. The reviewer ventures to express two hopes: (i) that nobody will use that "convenient unit" the millicentimetre^{-1} (p. 86)—what advantage has, say, "72 mcm.$^{-1}$" over the familiar "0·072 cm.$^{-1}$"? (ii) That a second edition will soon be called for, which will enable the author to correct a number of slips that have been found. It is usually stated that I is half-integer or zero according as the mass number is odd or even, except for ^{2}H and ^{14}N, which have I equal to 1. The author does well to emphasize that although this does describe the experimental results obtained up to the present, it should not be accepted as something in the nature of a *law*. It is true that lines of even-mass isotopes show no structure, and from this the spin is assumed to be zero in every case. But the nuclear masses for which the spin is known to be zero, from band-spectrum analysis, are not merely even but multiples of 4. So there is much to be said for the conclusion that I is half-integer or integer or zero according as the mass number is odd or an odd multiple of 2 or an even multiple of 2; to this there is no known exception.

W. J.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 47, PART 6 *November* 1, 1935 No. 263

536.721

THE EFFECT OF THE AERATION OF THE WATER USED IN THE DETERMINATION OF THE MECHANICAL EQUIVALENT OF HEAT

BY PROFESSOR T. H. LABY, M.A., SC.D., F.INST.P., F.R.S.
AND ASSOCIATE-PROFESSOR E. O. HERCUS, D.SC., F.INST.P.,
Natural Philosophy Laboratory, University of Melbourne

Received March 25, 1935. *Read in title June* 21, 1935

ABSTRACT. In relation to Dr Jessel's experiments two matters are discussed. (i) The effects of the liberation of dissolved air on the specific heat of water. It is shown that heat of solution, surface energy, and latent heat of vaporization contribute negligible amounts to the specific heat of water. A theory of the equilibrium of spherical air bubbles in fully aerated water is developed. (ii) The reliability of the existing evidence as to the value of the mechanical equivalent of heat. The conclusion is reached that Dr Jessel's experiments do not invalidate the work of Jaeger and Steinwehr or of the writers on the mechanical equivalent of heat, and that the value as calculated by Birge, viz. $4{\cdot}1852 \times 10^7$ erg/cal. at 15° C., is to be accepted.

§ 1. INTRODUCTION

Dr R. Jessel$^{(1)}$ finds experimentally that the specific heat of aerated water is greater than that of de-aerated water. He suggests reasons for this result, but does not discuss the theory exhaustively or quantitatively, and comes to certain conclusions as to the value at present to be accepted for the specific heat of water, i.e., the mechanical equivalent of heat. It is therefore desirable, as a guide to experiment, that the theory of the effect of aeration on the specific heat of water should be considered.

§ 2. THEORETICAL CALCULATION OF THE EFFECT OF AERATION ON SPECIFIC HEAT

As the following investigation shows, the calculated effect is very small. Aeration may affect the specific heat of water in three ways. (i) The heat of solution of the air. This has not been determined experimentally, though values have been obtained for some other gases. It can, however, be calculated from solubility data. If N is the mole fraction of the dissolved air, with an external pressure p, Henry's law may be expressed as N, p

$$k\mathrm{N} = p,$$ k

L and the heat of solution L cal./mole is given by⁽²⁾

$$\frac{L}{RT^2}=\frac{\partial \log_\epsilon k}{\partial T}.$$

Using the solubility tables of the *International Critical Tables*⁽³⁾, we find that

$$L_{15}=3349 \text{ cal./mole.}$$

This gives for the 0·42 cm.³ of air liberated from a litre of water per degree rise, 0·063 cal., and a consequent difference through aeration of the specific heat of water of the order of 1 in 20,000. This difference is not detectable by any experiment hitherto carried out.

(ii) The latent heat of the water vapour required to saturate the air bubbles. This will have its maximum effect on the specific heat when all the air liberated from solution by the rise of temperature of the water is retained in the form of bubbles in the water. Taking the case of 15° C.:

The volume of dissolved gas liberated per °C. is 0·42 cm.³ per litre of water⁽³⁾.
The mass of water vapour per cm.³ of saturated air at 15° C. is $1{\cdot}27\times10^{-5}$ gm.
The latent heat of vaporization at 15° C. is, say, 589 cal./gm.

This gives 0·0031 calories per litre, an increase of specific heat of the order 1 in 3×10^5.

P
p

(iii) The surface energy of the bubbles. The importance of this in its effect on the specific heat depends on the size of the bubbles, becoming greater as the average size of the bubbles becomes smaller. It is therefore necessary to obtain some evidence as to the relative stability of bubbles of different radii. Dr Jessel* gives $3a/2T$ for the smallest radius of a stable bubble; this appears to be a misquotation from Callendar⁽⁴⁾, who gives $\sqrt{(3a/2T)}$ as the greatest radius of a stable bubble for the case when P, the vapour pressure within the bubble, is greater than p, the external pressure. In the case considered here, in which $P<p$, Callendar shows that stable bubbles may have any radius. He is dealing, however, with a bubble containing a fixed mass of air; the following discussion allows for the passage of air from the bubble into solution or *vice versa*.

r, P_t
t
σ, d, T

The equilibrium of a spherical bubble in air-saturated water at uniform temperature. Let the external pressure on the free surface of the water be p dyne cm.⁻², and suppose the water to be saturated with air at this pressure. The vapour pressure P in a bubble of radius r cm. will be $P_t(1-\text{N})-\frac{2T}{r}\cdot\frac{\sigma}{d}$, where P_t is the saturation vapour pressure at temperature t as given in tables, N is the mole fraction of the dissolved air, σ, d are the densities of water vapour and water respectively, and T is the surface tension.

p'

For mechanical equilibrium, the air pressure p' within the bubble must satisfy the equation

$$p'+P_t(1-\text{N})-\frac{2T}{r}\cdot\frac{\sigma}{d}=p+dhg+\frac{2T}{r} \qquad \text{......(1)},$$

h

where h is the depth of the bubble below the free surface at pressure p.

---

* See reference (1), p. 758.

The solubility of the air enclosed in the bubble is increased by the curvature, and for equilibrium in regard to solution

$$p'=p-\frac{2T}{r}\frac{\rho}{d} \qquad \ldots\ldots(2),$$

where ρ is the air-density at pressure p.

From equations (1) and (2) the radius for equilibrium is given by

$$P_t(1-\mathrm{N})=dhg+\frac{2T}{r}\left(1+\frac{\rho}{d}+\frac{\sigma}{d}\right)$$

and

$$r=\frac{2T}{P_t(1-\mathrm{N})-dhg}\left(1+\frac{\rho}{d}+\frac{\sigma}{d}\right) \qquad \ldots\ldots(3).$$

This equilibrium is unstable*, smaller bubbles if in mechanical equilibrium according to equation (1) will lose air into the solution, and larger ones will gain air at the expense of the solution. But in either case the rate of change will increase the more r departs from its equilibrium value; so that bubbles of radii near the equilibrium value will predominate. The following table shows values of r in microns given by equation (3) at depths h below the free surface of fully aerated water.

h (cm.) / t (°C.)	0	0·5	1	2	5
15	87	90	92	98	121
20	63	64	65	69	79
30	34	34	35	36	39
50	11	11	11	11	12
100	1·2	1·2	1·2	1·2	1·2

Effect of surface energy on specific heat. The table above shows that bubbles formed in aerated water at 15° C. have a radius of about 9×10^{-3} cm. when in equilibrium; bubbles of radii near that value will change only slowly in size, so that we may assume that most of the bubbles present in water that is being heated will have radii of about this value. If the 0·42 cm.³ of air liberated per litre of solution per 1° C. rise remained in the form of n bubbles of this size, the surface energy would be given by $4\pi r^2 nT$ while $\frac{4}{3}\pi r^3 n=0{\cdot}42$. Substituting $T=74$, $r=9\times10^{-3}$, we obtain for the surface energy $1{\cdot}04\times10^4$ erg or 1 in 4×10^6 of the specific heat. This value may be exceeded through the temporary existence of smaller bubbles, but it leaves a wide margin before any effective contribution to the specific heat could be expected.

---

* *Bumping.* The effect of dissolved air in facilitating ebullition is made clear by the above discussion. If water contains sufficient air to saturate it at 100° C., a bubble as small as 1·2 μ will be in equilibrium with the water at 100° C., and water vapour can evaporate steadily into the bubble as the water is further heated, however slow may be the excess supply of heat over that required for losses. On the other hand a bubble of water vapour of this size near the surface of de-aerated water under normal pressure would satisfy the equation

$$P_t=p+2T/r=2p,$$

and consequently it would be in equilibrium with the surrounding water only if the latter were at about 116° C.

It will be seen that none of the effects considered is sufficiently large to account for the experimental results obtained by Dr Jessel, and that the difference observed by him on aeration of the water requires further investigation. We observed that if air were liberated the rate of flow of the water was irregular, and this disturbed that steadiness in thermal conditions which is essential for accuracy in continuous-flow calorimetry. It is to be expected that the electrically heated wire used by Dr Jessel would be even more disturbed.

§ 3. THE VALUE OF THE MECHANICAL EQUIVALENT OF HEAT

Dr Jessel concluded from his experiment that "if exact and reproducible results for the specific heat of water are to be obtained, air-free water must be used" in calorimetric experiments.

This conclusion, if true, means that de-aerated water must be used in a determination of the mechanical equivalent of heat. There is no modern* determination of that constant in which de-aerated water was used, but a consideration of the two available experiments does not support the conclusion that aerated and de-aerated water differ in specific heat.

Dr Jessel does not mention in his discussion the critical reviews which Birge(5), Henning and Jaeger(6) have made of the determinations of the mechanical equivalent of heat. Both of these reviews reject all the electrical determinations made previous to 1908. The most valuable of the older determinations is no doubt that of Barnes, but neither the standard of resistance nor that of e.m.f. which he† used was sufficiently accurately known for the calculation of a value of J comparable in accuracy with that obtainable from the data now available. Birge, after a very thorough analysis of the two determinations, arrives at the conclusion that the electrical determination by Jaeger and Steinwehr is in close agreement with the direct determination by the present authors‡. L. H. made 19 determinations of J between 15°·62 and 16°·74 C. and 4 at 20°·50 C. Birge points out that in a reduction of our observations which he has made, the mean residual error of the group of 19 observations is 0·0003, while that of the group of 4 is 0·0022. He adds that an observation with a residual from 7 to 10 times the mean residual should be rejected according to any of the well-known criteria of rejection, and on this ground he rejects the group of 4. He then obtains $J_{15} = 4{\cdot}18526 . 10^7$ erg. A reduction by Callendar's formula gives 4·18516, in good agreement, and we accept his exclusion of our summer experiments.

---

* The view of several authorities on the electrical units is that to be satisfactory a determination of the electrical equivalent of heat must have been made since 1908, when the London International Electrical Congress was held, for it is not possible to convert into ergs with sufficient accuracy the electrical measurements made prior to that date. Thus "modern" is used here to mean since 1908.

† Barnes's standard of resistance was the British Association ohm, the absolute value of which is uncertain. His standard of e.m.f. was an old type of Clark cell for which in 1902 he gave the value 1·43325; in 1909 he corrected this to 1·4335, while Callendar, on King's experiments, gave $1{\cdot}4341 . 10^8$ e.m.u. The resulting uncertainty in J is greater than 1 in 1000.

‡ J. S. will be used subsequently for the first of these pairs of authors, and L. H. for the second.

The present authors knew that this group of 4 experiments at 20°·5 C. gave a result for J_{15}, about 1 part in 2000 higher than the other 19 experiments, Callendar's variation of specific heat with temperature being assumed, but as the 4 experiments agreed well together they were not rejected.

Birge gives the following values as a result of his calculations:

J. S. (electrical determination) $J_{15}=4{\cdot}1850\pm0{\cdot}0009\times10^7$ erg/cal.

L. H. (direct determination) $J_{15}=4{\cdot}1853\pm0{\cdot}0008\times10^7$ erg/cal.

Jessel's observations give*

Jessel (de-aerated water) $J_{15}=4{\cdot}1852\times10^7$ erg/cal.

Jessel (ordinary distilled water) $J_{15}=4{\cdot}1905\times10^7$ erg/cal.

He states that his values are correct to 1 part in 2000, that is to 0·0021. It will be seen that his value for de-aerated water agrees with the values found by J. S. and L. H., while his value for aerated water is about 1 part in 1000 greater.

The above figures are evidence that the waters used by J. S. and L. H. have the same specific heat to an accuracy of possibly 1 part in 10,000, and the same as that of de-aerated water to an accuracy of 1 part in 2000. Now J. S. used 50,000 gm. of water in a calorimeter closed to prevent evaporation, and it is probable that this water was not sufficiently saturated with air for the latter to come out of solution when its temperature rose to the region of 15° C. Even if the water were saturated with air the vigorous stirring would prevent the formation of stable small bubbles, the effect of which has been discussed. We used ordinary distilled water under a pressure sufficiently above atmospheric to prevent air bubbles being formed as the water was heated in the 19 experiments at 16° C. The following is quoted from our paper(7) (p. 77) "*Steadiness of conditions.* The most serious temperature variations arise from variations in the rate of flow of the water. At first the flow was regulated by a carefully designed tap placed where the water enters the apparatus. The flow slowly decreased, and was subject to constant small irregularities, the cause of which was not at first evident to us, but later we found these fluctuations in the flow to arise from liberation in the calorimeter of dissolved air. The remedy, moving the tap to the exit side of the calorimeter, was simple enough. The change

---

* Jessel found the specific heats given in the following table:

J_t in 10^7 erg per calorie at t° C.

Ordinary distilled water			De-aerated water		
Temperature (°C.)	J_t	J_{15}	Temperature (°C.)	J_t	J_{15}
12·3	4·1944	4·1914	16·8	4·1858	4·1874
20·1	4·1857	4·1897	22·1	4·1780	4·1830
	Mean	4·1905		Mean	4·1852

His values have been corrected to 15° by means of an expression of the fourth degree in t calculated by Birge to fit J. S.'s values.

increased the pressure of the water from about 60 to 90 cm. of mercury, and at this higher pressure the water is unsaturated with air until after it has left the apparatus."

A study of the solubility tables will show that the excess pressure of nearly half an atmosphere in the calorimeter would compensate with a wide margin for the tendency of the rise of temperature of about 5° C. to liberate air. It is to be noted too that the water passed before entering the calorimeter through a filter of very fine metal gauze which would remove bubbles larger than the mesh of this gauze.

§ 4. CONCLUSION

Neither the theoretical discussion which has been given nor the experimental results of Jaeger and Steinwehr and of the present authors support the conclusion that aerated and de-aerated water have different specific heats. The value of the mechanical equivalent of heat for both de-aerated and aerated water in which the air remains in solution is $4{\cdot}1852 . 10^7$ erg per 15° C. calorie.

REFERENCES

(1) R. Jessel. *Proc. phys. Soc.* **46**, 747 (1934).
(2) Guggenheim. *Modern Thermodynamics*, p. 94 (1933).
(3) *International Critical Tables*, **3**, 258.
(4) *Encycl. Brit.* 11th ed. **27**, 898 (1911).
(5) Birge. *Rev. mod. Phys.* **1**, 1 (1929).
(6) Henning and Jaeger. *Handb. Phys.* **2**, 497 (1926).
Jaeger. *Handb. Phys.* **9**, 476 (1926).
(7) T. H. Laby and E. O. Hercus. *Philos. Trans.* **227**, 77 (1927).

DISCUSSION

Dr R. Jessel. The authors seek to justify their determination of the mechanical equivalent of heat, carried out by the use of water containing a certain amount of dissolved air, by showing that my tentatively suggested theory is not correct. This does not, however, alter the experimental fact, viz. that the presence of dissolved air increases the specific heat, a fact which has been proved both for crude oils and for petroleum fractions as well as for water. They only show that the theory should be modified to suit the facts.

The ideas expressed in figure 4 of my paper are still probably true. The contribution of the vapour molecules must be increased by the presence of air; i.e. the formation of steam molecules is greatly facilitated. This increase means that the air in the water in some way unknown helps to break down the highly coaggregated water molecules into simpler coaggregations or even into single molecules. Another important point which has not been mentioned is that for aerated water the minimum value occurs at a lower temperature. Barnes and I find it at about 37° C., Jaeger and Steinwehr at about 33° C. It appears from experiment that the *liberation* of air is not essential. Its presence seems to be all that matters.

In my paper I make no pretence at a standard determination of J_{15}, but to the accuracy of my experiment I find that the results agree with the corrected values of Barnes, who used de-aerated water, and it is his curve that is plotted. The suggestion made is that his corrected value of 4·1842 should be accepted and not the value 4·1860 or even the corrected value of 4·1853 of Laby and Hercus. Laby and Hercus's results at 20° C., which were self-consistent, seem to have been thrown overboard for the simple reason that the deduced value of J_{15} is too high. The presence of air seems an adequate reason for the high result.

I would suggest that the authors settle the matter for themselves by repeating their experiments, after modifying the apparatus if necessary, at 20° C. with air-free water; and also at higher temperatures, say 30° C. and 40° C. where the effect is much more noticeable, with both air-free water and aerated water. The results obtained by Bousfield, who also claimed a high degree of accuracy, seem to indicate that air has a considerable effect.

It is also suggested that the accuracy of the aerated-water experiments may be in some doubt, but this suggestion is entirely unfounded. Practically all the experiments were made with three or more rates of flow, so that the probable error is less than one part in 3000; moreover at temperatures below 35° C. the conditions were as steady as those for de-aerated water.

Dr H. R. Lang. If the authors examine the results of our experiments on twenty-four samples of oil and petroleum fractions they will surely agree that the increase in specific heat caused by dissolved gas is in general too large to be accounted for entirely in the manner they suggest. I would particularly refer them to the discussion of these results*. Dr Jessel's recent experiments with water confirm our conclusions. The satisfactory nature of his experiments is proved by the close agreement of his results for de-aerated water with those of Barnes, over the whole range from about 10° C. to 70° C. Using precisely the same method and apparatus he finds consistently higher values when the water is not de-aerated than when it is, a fact supported by Laby and Hercus's results at 20° C. The authors' calculations seem to indicate that the air must have some effect additional to those they consider, in order to account for the large difference between the specific heat of water before and after de-aeration at temperatures above 20° C., say, where the curves begin to separate; for here the magnitude of the difference is several times too large to be accounted for in the way they suggest. The peculiar shape of the curve found by Lüdin and the Bousfields at higher temperatures also requires explanation, and here again the magnitude of the difference between their results and those of Barnes is much greater, I believe, than can be accounted for by considering only the effects mentioned by the authors.

I hope it will be possible for the authors to carry out some further experiments in the neighbourhood of 40° C. or some higher temperature, in order to test their theory under conditions familiar to them.

---

* Lang and Jessel, *J. Inst. Petr. Techn.* **18**, 863 (1932) and Lang, *J. Inst. Petr. Techn.* **17**, 589 (1931)

AUTHORS' reply. As the quantities in question appear to have the same value to an order of accuracy of 0·1 per cent, it is necessary to define them carefully to prevent the confusion which is present in important writings on the subject. This is done in the following table.

Energy required to heat 1 *gram of distilled water from* 14·5 *to* 15·5° C.

s_1, s_1'	Water saturated with air, heated at a pressure of 1 atmosphere, part of dissolved air liberated:	s_1 erg	s_1' watt-sec.
s_2, s_2'	Water with air in solution but no air liberated on heating:	s_2 ,,	s_2' ,,
s_3, s_3'	De-aerated water:	s_3 ,,	s_3' ,,

We showed thermodynamically that $s_1 = s_3$ to an accuracy of 1 in 10,000, and calculation along the same lines from the solubility data shows that between 15° C. and 20° C., s_2 exceeds s_3 by only 1 in 100,000. Using our own measurements and those of Jaeger and Steinwehr, we conclude that $s_1 = s_2 = s_3 = 4{\cdot}1852 \times 10^7$ in terms of the erg.

Dr Jessel measured s_1' and s_3' and found that $s_1' = 1{\cdot}001_3 . s_3'$. In his discussion above he rejects our thermodynamic calculations, although he does not point out any error in them, and only offers in their place qualitative reasoning incompatible with their conclusion. He states that experiment shows that dissolved air increases the specific heat of water and that apparently the liberation of air is not essential to that increase. We deny that he is entitled to state anything about the specific heat s_2 of water in which air is present but is not liberated, for in all his published experiments he used either de-aerated water, or water from which part of the air was liberated as the temperature rose. Assuming he has proved experimentally (though we are unable to reconcile this with our calculation) that $s_1' = 1{\cdot}001_3 . s_3'$, that does not prove that s_2 is not equal to s_3 to an accuracy of 1 in 10,000, which is an important issue.

Dr Jessel reasserts that Barnes's absolute value of J is to be accepted, and ignores the criticism, made by every authority on electrical units who has discussed Barnes's experiment recently, that his electrical standards of resistance and e.m.f. are too much in doubt to make his result of value. We believe Jaeger and Steinwehr's value, as an *absolute* determination, to have about ten times the accuracy of Barnes's.

Dr Lang discusses whether $s_1' > s_3'$. He does not accept our thermodynamical calculation which gives $s_1' = s_3'$ to an accuracy higher than 1 in 10,000. He explains our disagreement with Dr Jessel's experimental result $s_1' = 1{\cdot}001_3$ by saying we must have left some effect out of account, but makes no suggestion as to what effect has been omitted. He considers the facts that (i) he and Dr Jessel found s_1' greater than s_3' for twenty-four oils, and (ii) Jessel's value of s_3' agrees closely with Barnes's value over the temperature range 10° to 70° C. to afford additional evidence that $s_1' > s_3'$ for water. Both these facts and our thermodynamic calculations can be reconciled if it is assumed that the liberation of air in Callendar and Barnes's calorimeter changes the temperature-distribution as compared with the distribution which obtains when no air is liberated, and that this leads to values of s_1' which are too high by about 1 in 1000.

Preliminary experiments in our determination of J showed that Callendar's elementary theory, which Jessel uses to reduce his observations, is very difficult to realize in practice. In order to avoid this difficulty, we abandoned Callendar's method of eliminating heat-loss and devised a new method in which the loss was 0·04 per cent, and could be determined certainly to ± 10 per cent of its value. In Jessel's experiments different calorimeters were used with the de-aerated and the aerated water. The heat-loss for the one used with de-aerated water was apparently 1 or 2 per cent, and for the other was higher. Our experience with this type of calorimetry makes us believe that an error of 1 in 1000 in the values of s_1' and s_3' might arise under such conditions.

The magnitude of the effect under discussion can also be calculated from the heat of solution as obtained from solubility data. The specific heats of air-free and aerated water may be compared by considering a cycle in which water is saturated with air at 20° C., the water is cooled to 15° C., the air is drawn from solution at that temperature, and air and air-free water are heated separately to 20° C., completing the cycle. The solubility data used above give $L_{20} = 3120$ cal./mol., with a consequent difference of specific heats over the range 15 to 20° C. of 6×10^{-6} cal./gm. So, allowing for uncertainties in the solubility data, we may say that the difference due to this cause is not more than 1 in 100,000.

621.326.4:535.242.49

THE COLOUR-MATCHING OF TUNGSTEN-FILAMENT LAMPS BY MEANS OF A SINGLE PHOTOCELL AND COLOUR FILTERS

By J. S. PRESTON, M.A., F.Inst.P., A.M.I.E.E.,
National Physical Laboratory

Communicated by H. F. Buckley, M.Sc., F.Inst.P., May 23, 1935

ABSTRACT. A method for colour-matching tungsten light-sources is described, in which one photocell and two coloured filters are employed. The ratio of the total transmissions of the two filters, as measured by the photocell, is used as an empirical measure of the colour temperature of the source, a tungsten-filament lamp. For certain lamps and photocell-filter systems the relation between this ratio and the lamp-voltage is approximately linear. An example of the use of the method in checking the calibration of standards of colour temperature is described, reference being made to a set of N.P.L. colour standards.

§ 1. INTRODUCTION

A METHOD of using two photocells of different colour characteristics for the colour-matching of tungsten-filament electric lamps has been described by G. T. Winch(1).

Its basis is as follows. The ratio between the energy radiated by the lamp-filament in one spectral region to that radiated in another spectral region depends on the colour temperature of the source. The spectral regions having been defined, the ratio may be used as an empirical measure of the colour temperature of the source. Two such regions are selected by the two photocells, and the colour temperature of the source then determines the ratio of the outputs of the two cells. For colour-matching in practice, the more sensitive cell is shuttered to give unit ratio when the cells are illuminated by the colour standard. The latter is then replaced by the test lamp, whose voltage is regulated until unit ratio is again indicated. By the use of a null method it is secured that attainment of unit ratio shall correspond to zero deflection of a sensitive galvanometer, and thus a high sensitivity is attained. This method entails the use of two photocells, which increases the cost of the apparatus above that of the simple photometer, and involves the use of a large diffusing window in the case of directional measurements, with some uncertainty as to the perfection of the diffusion unless the light-sources are of similar size and design.

With these disadvantages in view, and also as a matter of theoretical interest, another method, in which the same principle is applied in a different way, is described in this paper. It has the advantage that an ordinary photoelectric photometer, with one photocell, can be used with the simplest of additions.

§ 2. THE METHOD

The present method employs one photocell, of sufficiently extended colour-sensitivity, with two interchangeable colour filters to select the two regions of the spectrum. The ratio of the cell-outputs with the two filters successively in position can be used as a measure of the colour temperature of the light-source. In this case two separate readings are taken to determine the ratio, and a null method cannot be used; the procedure for colour-matching described below is adopted.

This method is a particular case of a general method. If between a light-source and a receiver—e.g. an eye or photoelectric cell—various coloured filters are successively interposed, the corresponding outputs of the receiver, together with knowledge of the properties of the filters, give information concerning the relations existing between the spectral distribution of the energy radiated by the source and the spectral sensitivity of the receiver. The amount of information thus obtainable depends upon the number and properties of the filters, and the data available concerning them. In particular if either the quality of the radiation or the properties of the receiver remain unchanged, variation in the factor that is subject to change can be deduced from the readings on the set of filters. For instance, if a source of constant colour be used, the colour characteristics of various human observers can be classified according to the ratio (the "Y/B ratio") which each obtains for the transmissions of two standard yellow and blue filters; or the performance of a photoelectric cell may be assessed by comparing the overall transmissions of a set of colour filters as measured with the cell with those determined on the visual basis, all the values being related to a source having a constant colour temperature(2). Alternatively if a receiver of constant properties, e.g. a photocell, is employed, the transmission readings for a set of filters will afford information as to the colour of the source. In the case of a tungsten-filament source, for which the energy-distribution bears a unique relation to the colour temperature, the transmission of a single filter will, in general, bear a unique relation to that colour temperature, and may be used as a measure of it. In certain cases the relation will not be single-valued; and if the filter be strictly neutral, its transmission will of course be invariable.

In the method herein described, however, two filters are employed for the practical reasons that a higher rate of change of the measured quantity with colour temperature of the source is obtained for the usual cells and filters available, and that if the filters are suitably chosen the photoelectric photometer may always be used at nearly its full scale-reading, with advantage in regard to precision.

The procedure for colour-matching two lamps is as follows. A photoelectric photometer is set up with the cell exposed to light from the colour standard, which is adjusted to the desired temperature. Two readings of the photometer are taken with, say, a red glass and a green glass in front of the cell. The green/red ratio is thus found. The colour standard is then replaced by the test lamp which is to be matched with it, and the procedure is repeated for two or more voltages on the test lamp. Appropriate interpolation gives the voltage on the test lamp required to give the same green/red ratio as the standard, i.e. to colour-match it.

§ 3. PRACTICAL CONSIDERATIONS

The results quoted below were obtained with an emission photocell in the well known thermionic-valve bridge circuit. The photometer need not be described here, but the following considerations assume the use of such a photometer, although many of them are of general application.

The photocell. The photocell should have a wide spectral sensitivity band and stable characteristics. A rubidium-on-silver thin-film vacuum cell with central plate cathode is a suitable type. The cell may have an illumination-response characteristic of one of three types, namely (i) strictly linear; (ii) not linear, but of similar form for all wave-lengths and such that the cell will always give the same measured value for the transmission of a filter, no matter what illumination-value of given colour temperature is used; (iii) not linear, and of different form for light of different wave-lengths.

In cases (i) and (ii) the procedure outlined above is valid. In case (iii), however, the cell will exhibit the Purkyňe effect, and the green/red ratio for a given colour temperature will vary according to the degree of illumination employed. It will then be necessary to work always with the same reading for one or other filter. This is best done by having the light-source movable on a photometer bench and always adjusting its distance from the diffusing cell window so that a definite constant reading is given by the photometer for one of the two filters. The reading on the other filter will then have one definite value for a definite colour temperature, and so will the ratio. In practice it is useful to set the reading for the less dense filter at the maximum scale-reading of the photometer.

A diffusing window in front of the photocell is a necessity, and is of particular importance in the case where the light-source is movable.

The filters. Filters suitable for use with the rubidium cell are a selenium red and a signal green glass. These should be optically finished, of good uniformity, and carried in a mechanical slide to ensure accurate placing and replacing in the path of the light. It may be necessary to combine the selenium glass with a neutral filter to give it an effective density more nearly equal to that of the green glass, so that both readings shall fall near the upper end of the photometer-scale, with consequent improvement in precision.

The main source of error likely to be encountered in respect of the filters is change in temperature, with a corresponding change in transmission. To keep the error small, a change of temperature exceeding 1° C. should not be tolerated.

§ 4. RESULTS

The following description of results obtained by the above method will illustrate its application.

(i) The method was used to investigate the mutual agreement, at a nominal colour temperature of 2360° K., of a set of six colour standards, and to readjust individual lamps to the mean of the group.

The lamps used were N.P.L. working standards 87G-L vacuum squirrel-cage lamps of nominal rating 100 V., 60 W. and had been calibrated visually. The set-up was as described above with a rubidium cell showing Purkyňe effect. The measured transmission of the red filter was reduced by cementing it to a neutral filter. The photometer was mounted on a bench and fitted with a diffusing window.

Lamp 87G was selected for use as a reference lamp against which to check the other five. Its stated voltage for the colour temperature of 2360° K. was 85·9. Its green/red ratio was measured at 85, 85·9 and 87 volts, the red reading being the greater throughout and so being set at a constant (full-scale) value. The relation between voltage and green/red ratio, figure 1, is seen to be a straight line over this range.

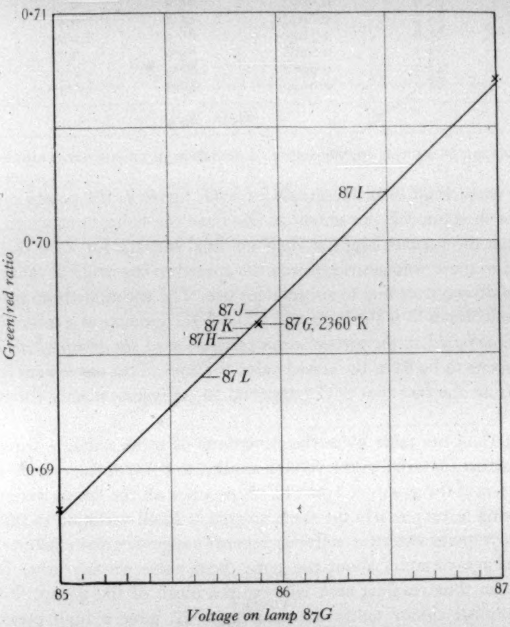


Figure 1.

The green/red ratio was then measured for each of the other five lamps, run respectively at the voltages stated in the calibration charts to give a colour temperature of 2360° K.; see the table, column A. If the stated voltages had been exactly the correct ones these five ratios should have agreed precisely with that for 87G at 85·9 volts. This agreement did not occur, the observed ratios being as in the table, column B.

Table.

Lamp	A Nominal voltage for 2360° K.	B G/R ratio at voltage in column A	C Voltage on 87G to give ratios in B	D* Deviation from mean of column C	E Corrected voltages for colour temperature of 2360° K.
87G	85·9	0·6963	85·9	0	85·9
87H	85·5	0·6957	85·8	−0·1	85·6
87I	85·6	0·7020	86·5	+0·6	85·0
87J	85·3	0·6968	85·9	0	85·3
87K	85·3	0·6961	85·9	0	85·3
87L	85·2	0·6940	85·6	−0·3	85·5
			Mean 85·9		

* A deviation of 0·1 volt corresponds to a deviation in colour temperature of 1° K.

If now we pick off from the graph for 87G, figure 1, the points corresponding to the ratios in column B, we obtain as abscissae the voltages at which 87G would colour-match the various lamps at their nominal settings for 2360° K., column C. The spread in these voltages represents the spread in the original calibration of the group, plus divergences due to subsequent use. The mean of these voltages, given at the foot of column C, is the voltage at which 87G operates at a colour temperature of 2360° K. *as based on the present mean calibration of the group of six lamps.* This voltage happens to be 85·9, the stated value for 87G. This agreement is fortuitous, and arose from the fact that 87G happened to represent exactly the mean of the group.

Column D of the table gives the deviations of these voltages from the mean, i.e. the departure, in terms of the voltage applied to 87G, of the colour of each lamp from the mean of the group at 2360° K. Now since all the lamps were of the same type, operating at very nearly the same voltage, a small variation in the voltage on 87G will correspond almost exactly, as regards range of colour temperature, with an equal voltage-variation about the same fixed value on any other lamp in the group. We can thus readjust each lamp to the mean of the group. For instance, 87I at a nominal colour temperature of 2360° K. gave a high green/red ratio, figure 1, indicating that the stated voltage was too high by an amount equivalent to an increase of 0·6 volts on 87G. The stated voltage of 87I should therefore be reduced by a like amount. By such a process we arrive at the values in column E, which are the voltages at which all six lamps accurately match in colour, and at which they represent the mean 2360°-K. point for the group, *on the basis of the present calibration.*

A second complete redetermination of the values in column E led to the same figures within 0·1 volt in every case, showing that the precision of the method corresponds to a colour-temperature range of about ± 1° K. at 2360° K.

(ii) The form of the relation between green/red ratio and lamp voltage was found in case (i) to be linear over a range of 2 volts for 87G. As a matter of interest this range was extended to 40 volts, and the measurements of the ratio were made at intervals of 5 volts with two entirely different cell-filter systems. The results are shown in figure 2. Curve A was obtained with a rubidium cell which showed the Purkyňe effect. The filters were a selenium red glass, attached to a neutral filter, and a signal green glass. The conditions were not precisely as in case (i).

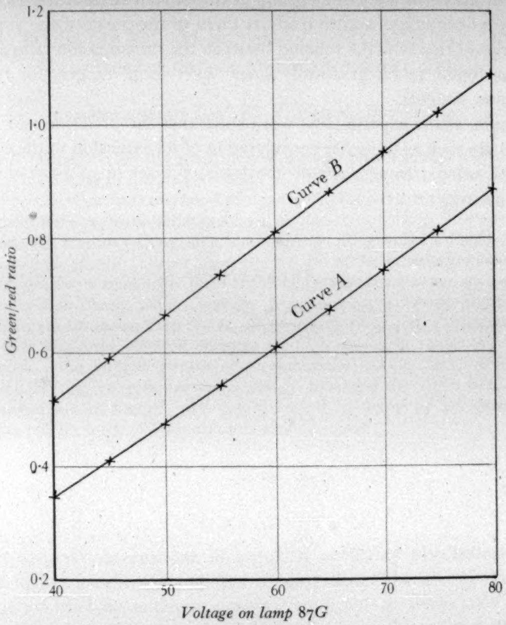


Figure 2.

Curve B was obtained with a thin-film potassium-on-silver cell. The filters in this case were a dark-orange-plus-orthogreen combination, and a signal green glass. The relations are not precisely linear, but so nearly so that linear interpolation over ranges of a few volts does not involve errors as large as the experimental errors. In colour-matching a pair of lamps, therefore, the green/red ratio for the standard having been determined, measurement of the ratio for two voltages only on the test lamp gives sufficient data to enable the colour-match voltage to be found by interpolation, provided it is close to the measured points.

§ 5. CONCLUSIONS

The photoelectric determination of the ratio of the transmissions of two coloured filters gives a convenient and precise method for colour-matching tungsten-filament electric lamps. The method has been successfully used to ascertain the spread in the performance of a group of colour standards of the vacuum 100-V., 60-W. squirrel-cage type, and to readjust them to the mean of the group.

For lamps of this type the relation between the transmission ratio and the lamp voltage was found to be practically linear over large ranges, for two different photocell-filter systems.

The results of the experimental work show that the precision and reliability of the method are such as to justify consideration of its extension to the establishment of a scale of colour temperature on the basis of a set of calibration points independently derived.

REFERENCES

(1) WINCH, G. T. *J. sci. Instrum.* **6**, 374 (1929).
(2) PRESTON, J. S. and McDERMOTT, L. H. *J. sci. Instrum.* **11**, 153 (1934).

537·525   681.2.087.47

SPECTROSCOPIC OBSERVATION OF RECURRENT PHENOMENA IN DISCHARGE TUBES

By R. H. SLOANE, M.Sc. and C. M. MINNIS, M.Sc.,*
Department of Physics, Queen's University, Belfast

Communicated by Prof. W. B. Morton, May 21, 1935

ABSTRACT. A method for the synchronous photography of irregular moving striations is described. It consists in forming an image of the positive column of a discharge along the slit of a spectrograph and covering and uncovering the latter with a mechanical shutter made to synchronize with the striations by a photoelectric device. The spectra obtained show variations in intensity along each line, corresponding to the intensity of the line in the bright and dark phases. Direct photography also of the actual striations is described.

The photographs obtained for pure argon show that there is no appreciable recombination in the dark phase, while those for argon-mercury mixtures show excitation of only the mercury lines in the dark phase. These results have been explained by assuming that in the dark phase the electron energies are high enough to excite the mercury lines and are therefore too high to permit of recombination taking place. It is only in the bright phase that the energies are large enough to excite the argon lines. The optical results described are in accord with data obtained by Pupp by an electrical method. Their bearing on the work of Ladenburg is also discussed.

§ 1. INTRODUCTION

Stroboscopic observations of regularly occurring phenomena are usually made with a synchronous slotted rotating disc. The aim of the present investigation has been to devise a method applicable to phenomena which do not occur with constant frequency and to which, therefore, the ordinary method is not applicable. This has been achieved by arranging to have a mechanical shutter opened and shut in synchronism with, and directly controlled by, the individual events. Although primarily designed for an investigation of moving striations in low-pressure discharges, the method is general and could have wider applications than those indicated here.

Briefly, the method consists in applying (for example by means of a photocell) voltage impulses caused by the individual events to a thermionic valve amplifier which feeds an output stage capable of giving a large current; this current is passed through a modified form of Einthoven string galvanometer, so that when the fibre is deflected it uncovers a slit through which direct observations of the events are made.

---

[*] Musgrave Student in Physics.

§ 2. CIRCUIT

The circuit is shown in detail in figure 1. Light from the discharge tube *T*, in which moving striations were present, was admitted to a vacuum photocell *P* through a fine slit in an earthed metal box. The slit could be either vertical or horizontal and its dimensions could be varied to control the intensity of the light incident on the photocell. The current through the photocell developed a potential difference across a resistance R_1 in such a way that each striation, as it passed the photocell, gave rise to a negative voltage impulse which was applied to the control grid of a screen-grid valve V_1 with variable amplification factor.* The output from the latter was passed into a three valve resistance-capacity coupled amplifier *RCA* of conventional design, which in turn fed an amplifier, consisting of two pentodes† in parallel, through a transformer and a potentiometer device R_2 which regulated the input to the pentodes. The anode current of the pentodes operated the electromagnetic shutter.

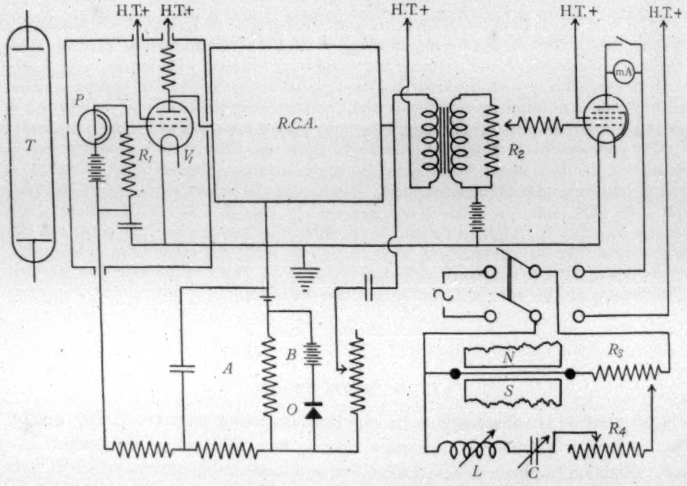


Figure 1. Circuit diagram.

Since it was desirable to have the current in the shutter unidirectional, the two pentodes were used with a large steady negative voltage applied to their control grids so that the anode current could only flow during the positive half-cycle of the input voltage from the previous amplifier. When the shutter was being operated, the average anode current to these valves, measured on a moving-coil milliameter mA, was adjusted to be between 70 and 90 mA., varying in extreme cases by no more than 5 per cent, while the quiescent current never exceeded 3 mA.

---

* Osram VS 24.

† Osram PT 16.

It was often necessary when photographic records were being made to leave the apparatus working continuously for several hours, and during this time the intensity of the light from the discharge tube was liable to vary considerably. The special function of the valve with variable amplification factor was to keep the output current constant in spite of such changes. This was accomplished by having its control grid connected to the output of the three-valve amplifier by a rectifying circuit *A*. This well-known circuit was so adjusted that when the output from the amplifier exceeded a certain voltage determined by a battery *B*, the control grid voltage of V_1 began automatically to be made more negative by the voltage developed across a copper oxide rectifier *O*, and hence the amplification produced was decreased. This system was more flexible than any utilizing the rectifying properties of the pentodes.

The pentode output current was adjusted initially to the required value by varying the input to the pentodes by a potentiometer device R_2, which while it controlled the amplitude had a negligible effect on the phase.

§ 3. CONSTRUCTION AND OPERATION OF SHUTTER

The design of the magnet and shutter is shown in figure 2; it applies the principle employed in the Einthoven string galvanometer. The pole pieces were bored as shown and were mounted at the centre of two end plates which were held together by the bobbin cores. The two field coils were wound for operation on the 220-volt d.-c. mains and were cooled by passing water through two spirals of copper tubing which formed the ends of the former of the coil, as shown. The inner ends of these spirals were soldered into a brass cylinder which was spaced from the iron core by a watertight brass ring at either end, permitting water to flow from one spiral to the other. Efficient water cooling was necessary to ensure constant temperature of the fibre mechanism during operation. As a further precaution, a small shield of copper foil was placed over each winding and soldered to the copper spirals on the side of the coil adjacent to the fibre.

The width and horizontal position of the spectrograph slit could be varied by means of thumb-screws *A* and *B*; these controlled two long levers, not shown, which in turn operated the slit. The fibre was stretched over two ebonite bridges *C* mounted on a brass carrier which could be withdrawn by slackening two thumb-screws *D*. Connexions were taken from the ends of the fibre to the terminals *E*. When the thumb-screw *F* was slackened, the rod *G* could slide freely in the ebonite terminal block, allowing the tension of the fibre to be adjusted as described in § 5.

The fibre could be set accurately parallel to the jaws of the slit by rotating its holder about a pivot *K* by means of the screws *L*. Horizontal movement of the fibre holder was effected by means of bevel gears *H*, which in turn were controlled from the back of the spectrograph by a long rod with a universal coupling.

Since oil damping of the fibre would have been impracticable, the motion was

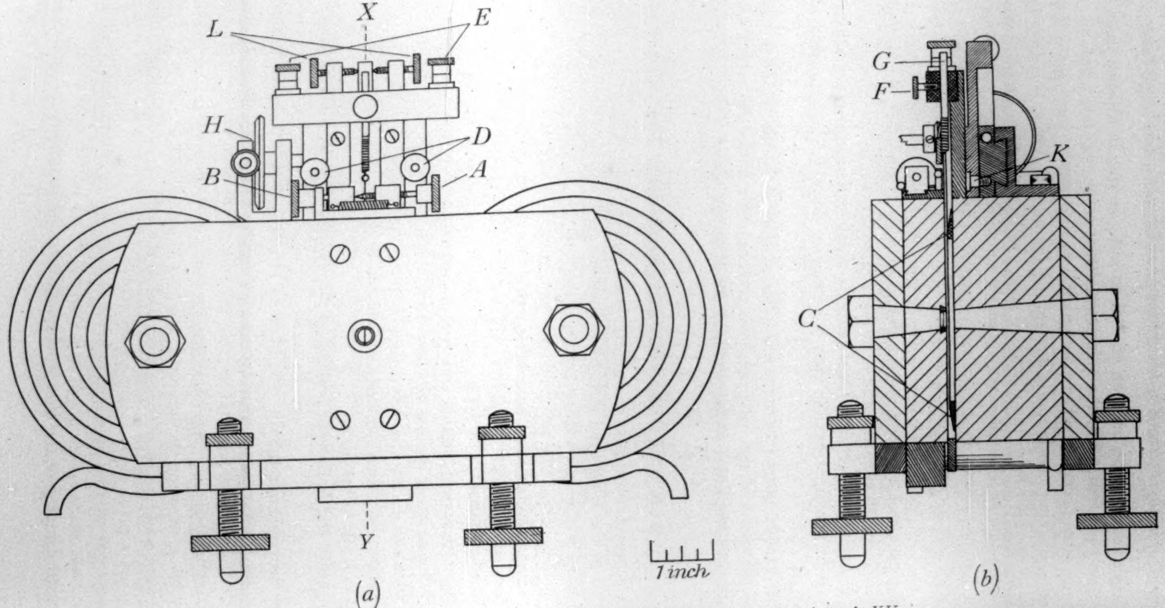

X
L
E
H
D
B
A
Y
1 inch
(a)
G
F
K
C
(b)

Figure 2. (*a*) Elevation of electromagnetic shutter. (*b*) Sectional elevation through *XY*.

damped electrically by an Irwin resonant shunt(1). In this method the fibre is shunted by a series tuned circuit *LC*, figure 1, with a frequency closely equal to that of the fundamental mode of the fibre. By means of the auxiliary resistances R_3, R_4, the amount of damping could be varied within wide limits. This system worked admirably.

§ 4. CONSTRUCTION OF DISCHARGE TUBES

Tube 1, figure 3, was used for discharges through argon-mercury mixtures and had nickel electrodes in the form of hollow cylinders, mounted on valve pinches. It was permanently fused to the pumping train and was baked for several hours under high vacuum before use, after which the electrodes were degassed with an eddy-current heater. Since only very small traces of mercury vapour were required, the tube was isolated from the mercury diffusion pump by a liquid air trap. Mercury vapour was admitted from a side-tube, which contained a drop of distilled mercury, by opening a tap for a few minutes at regular intervals. The tap

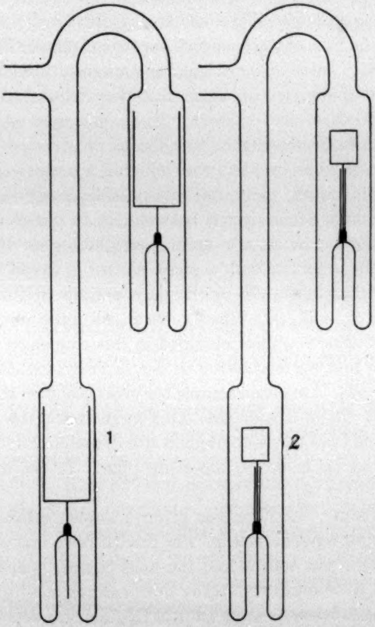


Figure 3. Discharge tubes.

was lubricated with apiezon grease, and apart from mercury lines the spectrum was entirely that of argon.

Tube 2 was intended for work on pure argon and on the auroral line of oxygen. It was similar in shape to tube 1, but had thick degassed tungsten plates for electrodes. Before the admission of argon the gas was first purified by means of a misch-metal arc, and in tube 2 a final treatment was given in the tube itself by heavy sputtering of the tungsten.

Tube 3 was a straight commercial tube, 50 cm. in length and 1·5 cm. in bore. It contained neon, and had hollow electrodes of sand-blasted steel in short T-pieces at the ends.

§ 5. APPLICATIONS

Spectra of moving striations. The magnet described in § 3 was designed to rest on an extension of the base of a glass-prism spectrograph, so that the electromagnet slit was almost in the position of the original spectrograph slit. By means of thumbscrews the position of the electromagnet was then adjusted so that the slit was accurately at the focus of the collimator lens. V-shaped stops into which the levelling-screws of the electromagnet fitted were provided, allowing the electromagnet to be removed when required and replaced accurately in position.

For spectrographic work, there are two possible positions for the discharge tube. (i) The tube may be connected horizontally to the pumping train and an image of it focused on the slit so as to cross it at right angles. By suitably adjusting the phase-shift in the amplifier with a phase-shifter it would then be possible to obtain spectra of either the bright or the dark periods of the moving striations. This method was not used. (ii) When vertical, the tube may be focused longitudinally on the slit. Spectrograms obtained in this way show regular variation in intensity along each line, corresponding to the varying intensity of the line in the bright and dark periods. A typical example for neon is shown in figure 4; the broad lines are due to the use of a wide slit. This method has the advantage that the spectra of both periods can be photographed simultaneously and direct comparison of the intensities can be made on the same plate. It has therefore been used exclusively.

Before each exposure, the fibre was given a known tension by inverting the holder and suspending weights from it. The loaded fibre was then replaced in the shutter, the water flow was started and the field current was switched on. After allowance of about 30 minutes for steady-temperature conditions to be attained, the fibre was allowed to resonate at its natural frequency, a separate valve oscillator being used as the source of current. With the phosphor-bronze strip, 6·3 cm. long by 0·17 mm. wide, under a tension of 70 gm. weight, used for most of the work, the frequency was about 1700 sec.$^{-1}$ The Irwin shunt was then connected and the inductance was varied until the vibration of the fibre, as seen in a microscope, was completely suppressed. After this, the fibre was oriented until it was accurately parallel to the jaws of the slit. The width of the slit was always adjusted with the magnet in position by observing the spectrum of the discharge in the back of the spectrograph. Then, with only the quiescent pentode current flowing, the fibre was moved over so as just to cover the slit. Finally the photocell was placed close to the discharge and exposed to it, when the striations could be seen in the back of the spectrograph in the various monochromatic wave-lengths emitted by the tube.

Direct photography of the striations. The apparatus was also used to obtain photographs of moving striations in the positive column of a discharge in neon. Tube 3 was first of all focused on the ground-glass screen of a camera. After the adjustment described in the previous paragraph, the shutter was interposed between the camera and the tube, so that only light which had passed through the shutter from the tube could reach the camera. Figure 5 was obtained in this manner with an exposure of 20 minutes on a fast panchromatic plate, and represents about 9 cm. of the tube. A photograph of this length is obtainable only when the velocity of the striations is constant throughout the region photographed. The dark line at X is due to a small appendix on the discharge tube.

Visual observations of the positive column made with tube 2 in pure argon showed that although the contours of the individual striations were very similar to those of figure 5, they were more widely spaced from each other. It should be noticed that the striations are actually less elongated than appears from such photographs, since they move during the time that the shutter is open. The fractional distortion depends on the conditions of adjustment. Its chief drawback is that it prevents us from observing accurately how the spectrum varies just at the edge of the bright part of the striation. The spectra of the main bright and dark periods can be recorded faithfully.

§ 6. DISCUSSION OF OBSERVATIONS

By means of the procedure described in § 5, spectrograms were obtained with tube 2 containing pure argon at pressures of 2–3 mm. The tube was run at 740 V. with currents of 40–80 mA. A careful examination was made of the spectra of the dark periods but there was no visible trace of high series lines or series continua, such as might be expected if recombination took place in that region.

It has been shown by Kenty⁽²⁾ that recombination of electrons with positive ions of argon in the gas phase is negligible when the electron temperatures exceed about 0·5 e-V. Hence we may conclude that the electron temperatures in the dark phase exceed this value in our tubes. This conclusion is in accord with the work of Pupp⁽³⁾, who by an oscillographic method has made probe measurements in moving striations in the positive columns of discharges through argon. Since the current-densities were of the same order,* in this work and in our own, we are probably justified in assuming that Pupp's results can be applied with fair accuracy to our discharges in pure argon. The results obtained in his tubes show that the minimum value of the electron temperature in any phase of the striation is not less than 1·5 V. Energies of this value preclude any possibility of appreciable recombination. Pupp's work also shows that the values obtained for ion-concentration repeat themselves very closely through several striations. Since the ions do not recombine in the gas, it follows that the average rate at which they are generated at any part of the positive column is equal to the rate at which they are lost from it (ultimately to the walls) under the action of the electric fields and concentration-gradients.

---

* The range of Pupp's current densities was 0·05–0·10 A.cm.$^{-2}$, while ours was 0·03–0·06 A.cm.$^{-2}$

Curve (i), figure 6, gives the number of ions generated sec.^{-1} cm.^{-1} along the axis of the positive column. It was obtained by substituting the numerical values for electron-concentration and temperature as given by Pupp,* in the kinetic-theory expression of Killian(4) for the rate of ionization by collision. In order to calculate the rate of diffusion to the walls, it would be necessary to consider an expression of the form(5)

$$D\left[\frac{e}{kT}\frac{dV}{dr}+\frac{1}{n}\frac{dn}{dr}\right],$$

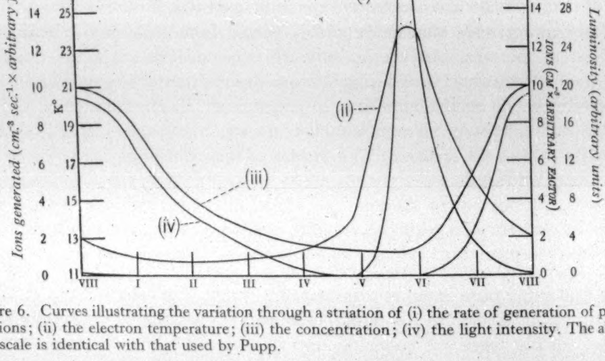


Figure 6. Curves illustrating the variation through a striation of (i) the rate of generation of positive ions; (ii) the electron temperature; (iii) the concentration; (iv) the light intensity. The abscissa scale is identical with that used by Pupp.

where D is the diffusion coefficient of the ions. Now the value of D is extremely sensitive to such factors as purity of the gas and temperature(6), and the conditions are unfortunately not well enough known to enable a reliable value of D to be estimated. Thus it is not at present feasible to make an exact estimate of the rate of diffusion of ions to the walls, but it is certain, from the optical indications, that were such a calculation possible it would be found that both the ionization and diffusion curves, when integrated over a complete striation, would give the same value. In other words just as many ions pass to the walls as are generated.

Figure 7 was obtained with tube 1 filled with argon at a pressure of 2 mm., and just sufficient mercury vapour to give the lines of both elements with comparable intensities—probably less than 0·03 per cent. This photograph, again obtained as described in § 5, shows the spectrum of a dark period along the centre and that of a bright striation on either side. The outstanding feature is the persistence of the mercury lines λ 5461 and λ 4358 through all phases, although argon lines of equal and greater intensity in the bright phases are relatively very weak in the dark phase(7).

---

* We assume, as usual, that the electrons have a full Maxwellian distribution.

The question arises as to the cause of this selective excitation. In the dark phase the excitation of the mercury lines could be due either to the persistence of excited systems from the bright period, or to direct excitation by direct electron collision in the dark period. If persistent excited systems were to be important, they would have to possess a lifetime of more than about $5 \cdot 10^{-4}$ sec., since the frequency of the striations is of the order of 10^3 sec^{-1} The persistence of ordinary excited states for this time is obviously out of the question, and even if either of the metastable argon atoms, with excitation potentials of 11·50 and 11·67 V., were present in the dark period, the probability of their exciting the mercury atom to the 7^3S_1 state, with an excitation potential of 7·6 V. by a collision of the second kind, with consequent emission of λ 5461 and λ 4358, would be negligible owing to the large energy-difference involved. There is a slight possibility that metastable mercury atoms, if present, could be excited to the 7^3S_1 state by electron impact, since Pupp's data for electron temperatures show that energetically this is possible.* However, apart from energy considerations, the concentration of these metastable atoms is hardly likely to be sufficient to account for the intensities of the observed lines, since a cumulative process is involved. The main excitation process then appears to be by direct collision of electrons with neutral mercury atoms.†

Curves (ii), (iii) and (iv) in figure 6 were derived from diagrams given by Pupp, and represent the variation of electron temperature, concentration and light-intensity through the striations.‡ They show that the region of maximum electron temperature precedes the front boundary of the bright striations, a result which appears at present to be inexplicable. We are therefore forced to the conclusion that either this result is spurious, or some fundamental fact bearing on the mechanism of moving striations still remains to be discovered. However, Pupp's results definitely show that throughout a cycle sufficient electrons are present with energies large enough to excite the mercury lines observed, although it is only in certain regions that strong excitation of the argon is possible.

It must be emphasized that precautions must be taken in work on intensities in spectra to ensure that the positive column shall be uniform, although this has not always been done. If the positive column is not uniform, great care is needed in interpreting the results obtained, for the presence of moving striations is liable to lead to large errors. For instance, the fact that large concentrations of atoms in many of the excited states are present only in the bright region indicates that in transverse absorption experiments, without synchronization of the source of light with the striations, one cannot hope to get more than about 50 per cent absorption when these states are involved. It is only when metastable atoms with fairly long lifetimes are responsible for the absorption that it will be appreciable in all phases, since the long lifetimes of these states will tend to maintain a reasonably high concentration through the striation. When the emission spectrum of a discharge showing moving striations of this type is examined laterally without any stroboscopic device, it is also obvious that the spectrum obtained may show an anomalous distribution of intensities among the observed lines due to the preferential emission in the dark period of some lines of the type which we have found. Finally, in the case of measurements of longitudinal absorption, as in the work of Ladenburg⁽⁸⁾, the same kind of averaging effect is present. It is only to be expected, in accordance with his results, that electron temperatures deduced from data obtained in this way will not give anything better than order-of-magnitude agreement with results obtained by electrical methods, such as those of Seeliger and Hirchert⁽⁹⁾, in which the probe is used continuously, and not intermittently in synchronism with the striations if such are present. Used intermittently, the probe will give true temperatures for the particular phase of the striation during which it is in circuit⁽²,³⁾, but used continuously it will merely give an average temperature which is not necessarily the same average as that yielded by the optical method.

---

* Since in the Maxwellian distribution there will be an appreciable number of electrons with energies twice that of the mean value.

† Although Pupp's data refer to discharges through pure argon, we can probably apply them to our case without appreciable error.

‡ The curves shown refer to the axis of the tube.

§ 7. CONCLUSION

It is possible that the circuits employed here could be improved in detail; those used were found quite satisfactory for this investigation and permitted the use of apparatus already available. With modern high-efficiency valves, adequate amplification could be obtained with fewer stages. Further development could apparently be effected by application of the method to the observation of plasma oscillations⁽¹⁰⁾, by modifying the circuits and using a Kerr cell or other shutter suitable for very high frequencies.

The results for discharges through argon-oxygen mixtures will be discussed in a forthcoming publication.

§ 8. ACKNOWLEDGMENTS

We wish to thank Prof. Emeléus for suggesting the problem and providing facilities for the experimental work, and for criticism and advice throughout it. Our thanks are also due to Mr J. Wylie, B.A., for valuable help with the design of the magnet which was constructed in the Physics Workshop by Mr A. Buick. The commercial neon tube was kindly lent by the General Electric Company through Mr Morrison, of the Belfast branch.

REFERENCES

(1) Irwin, T. J. *Oscillographs*, p. 91 (Pitman, 1928).
(2) Kenty, C. *Phys. Rev.* **32**, 624 (1928).
(3) Pupp, W. *Phys. Z.* **36**, 61 (1935).
(4) Killian, T. J. *Phys. Rev.* **35**, 1238 (1930).
(5) Darrow, K. K. *Electrical Phenomena in Gases*, p. 185.
(6) Tyndall, A. M. and Pearce, A. F. *Proc. roy. Soc.* A, **149**, 426 (1935).
(7) Sloane, R. H. and Minnis, C. M. *Nature*, Lond., **135**, 436 (1935).
(8) Ladenburg, R. *Rev. mod. Phys.* **5**, 243 (1933).
(9) Seeliger, R. and Hirchert, R. *Ann. Phys.*, Lpz., **11**, 817 (1931).
(10) Revans, R. W. *Phys. Rev.* **44**, 798 (1933).

538.214:546.711

THE MAGNETIC MOMENT OF THE MANGANIC ION

By L. C. JACKSON, M.Sc., Ph.D., Wills Physical Laboratory,
University of Bristol

Received June 9, 1935

ABSTRACT. The magnetic susceptibility of manganic acetylacetone has been measured for a powdered specimen from 292° to 16·9° K. The law $\chi(T+5{\cdot}5)=$ a constant is obeyed down to about 75° K., below which the susceptibility increases somewhat more rapidly than is indicated by this law. The magneton number for the Mn^{+++} ion, for which no reliable data were previously available, is 4·98 as calculated from the above relation. This is in good agreement with the requirements of the Bose-Stoner theory.

§ 1. INTRODUCTION

The available data on the magnetic properties of substances containing the manganic ion Mn^{+++} are very meagre on account of the fact that these compounds are generally very unstable. They are practically restricted to Weber's(1) measurements on solutions of certain manganic salts and Honda and Soné's(2) on manganic oxide Mn_2O_3. The former measurements were made at room-temperature only and, on the assumption of the validity of Curie's law, the most reliable data, those on manganic ortho-phosphate, $MnPO_4$, lead to a magneton number of 5·03. On the other hand, when Honda and Soné's values of the magnetic susceptibility χ are plotted on the customary $\{1/\chi, T\}$ diagram, it is seen that the curve is concave to the temperature axis over the whole range of the observations, and therefore only an effective magneton number which varies considerably with the absolute temperature T can be calculated.

χ

T

The metallic acetylacetones of the general formula $M(CH_3.CO.CH.COCH_3)_3$, in which M is any tervalent metal, form a series of stable readily crystallizable compounds. Although these compounds cannot be regarded as simple salts, on account of their non-ionizability and their solubility in organic solvents, the metallic ion seems frequently to possess the same ground state as in the corresponding simple salts. Thus ferric acetylacetone(3) has the same magneton number 5·94 as the simple ferric salts, and it seemed worth while to investigate the magnetic properties of manganic acetylacetone over a wide range of temperature in the hope of obtaining a reliable estimate of the magnetic moment of the Mn^{+++} ion.

§ 2. METHOD AND MATERIAL

The susceptibilities were measured by the Faraday method with the aid of an apparatus fully described in a previous paper(3). The specimen was sealed in a small spherical phial of pyrex glass and occupied the place of the crystal and carrier shown in figure 1 of the paper in question. The low temperatures were obtained by means of baths containing liquid ethylene, oxygen, nitrogen and hydrogen boiling under atmospheric or reduced pressure, and of a mixture of solid carbon dioxide and acetone. The temperatures were determined by means of a copper-constantan thermocouple, or in the case of liquid oxygen and hydrogen, which were adequately pure, from the vapour pressure of the liquid in the bath.

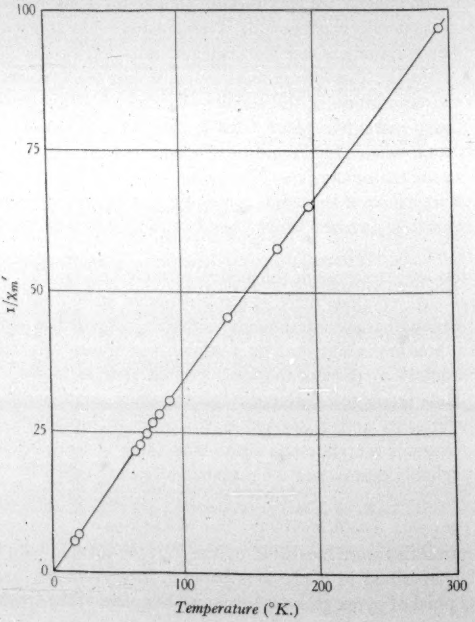
The manganic acetylacetone was kindly prepared for me by Dr R. C. Menzies of the Chemistry Department of Bristol University. It was further recrystallized from acetone and a phial filled with the finely ground material. The substance was apparently in its monoclinic form(4) but this observation was not checked by actual goniometric measurements.

§ 3. EXPERIMENTAL RESULTS

The specific susceptibility of manganic acetylacetone was measured from room-temperature down to 16°·95 K. with the results given in the accompanying table.

χ_m'

The reciprocal of the molecular susceptibility χ_m', corrected for the diamagnetism of the organic groups in the molecule, is plotted against the absolute temperature in the figure. It will be seen that the substance follows the Weiss law $\chi\,(T+\Delta)=a$



constant with Δ equal to $+5\cdot5$ down to about 70° K. Small deviations from this law occur at lower temperatures, the susceptibility increasing more rapidly with T than is given by $\chi\,(T+5\cdot5)=$ a constant.

Table. Manganic acetylacetone

T (° K.)	$\chi\times10^6$	χ_m'	$1/\chi_m'$
292·2	$28\cdot8_8$	$0\cdot0103_5$	$96\cdot6_6$
196·0	$43\cdot1_0$	$0\cdot0153_6$	$65\cdot1_0$
172·7	$48\cdot1_8$	$0\cdot0171_4$	$58\cdot3_4$
135·2	$61\cdot9_0$	$0\cdot0219_7$	$45\cdot5_2$
90·2	$92\cdot1_5$	$0\cdot0326_3$	$30\cdot6_6$
82·3	100·0	$0\cdot0353_9$	$28\cdot2_6$
77·4	105·4	$0\cdot0372_9$	$26\cdot8_2$
72·7	113·6	$0\cdot0401_8$	$24\cdot8_9$
67·4	123·5	$0\cdot0436_7$	$22\cdot9_0$
63·9	129·6	$0\cdot0458_1$	$21\cdot8_3$
20·3	425·5	0·1500	6·66
$16\cdot9_5$	510·5	0·1799	5·56

§ 4. DISCUSSION OF RESULTS

The magneton number calculated in the customary way from the straight part of the curve is 4·98 Bohr magnetons. It is thus quite comparable with that given by Weber's work on solutions.

According to the theory of Bose-Stoner as worked out in detail later by Van Vleck the magneton numbers of the ions of the elements of the first transition group should lie between $\sqrt{\{4S\,(S+1)\}}$ and $\sqrt{\{4S\,(S+1)+L\,(L+1)\}}$, that is to say between the value for spin only and that for spin and orbital moment both effective but uncoupled from one another. In the first half of the group the L contribution is almost completely suppressed and the observed magneton numbers agree closely with the spin values. For the Mn^{+++} ion, the ground state of which is $^5D_0$ from Hund's rules, the magneton number should lie between 4·90 and 5·48 but approximate closely to the first of these values. It will be seen that this is the case for the value deduced above from manganic acetylacetone.

REFERENCES

(1) Weber. *Ann. Phys.* Lpz., **19**, 1056 (1906); **36**, 624 (1911).
(2) Honda and Soné. *Sci. Rep. Tohoku*, **3**, 139 (1914).
(3) Welo. *Phil. Mag.* **6**, 481 (1928).
Jackson. *Proc. roy. Soc.* A, **140**, 695 (1933).
(4) Astbury. *Proc. roy. Soc.* A, **112**, 448 (1926).

535·23:535·241.3

THE COLOUR TEMPERATURES OF THE HEFNER AND ACETYLENE FLAMES

By A. R. PEARSON, M.Sc., LL.B. and B. PLEASANCE

Received April 13, 1935

ABSTRACT. The colour of the "cylindrical" acetylene flame depends on three factors, namely the gas-pressure, the rate of consumption and the height of the flame. Discordant statements in the literature are believed to be due to lack of uniformity in these particulars. The flame specified by Dr Jones of the Kodak laboratory is 50 mm. high and consumes 0·25 ft.^{3}/hr. at 90 mm. (water) pressure. A 1-cm. section of a flame conforming closely to this specification, burning on a Bray Vika tip, is found to have colour temperature 2520° K. The flame on a Bray Ota tip at 90 mm. pressure is 58–60 mm. high and consumes 0·29–0·30 ft.^{3}/hr. One cm. sections of two such flames were found to have colour temperatures (in round figures) 2380° and 2400° K. We have not been able to test a large enough number of burners to say what limits of variation may be expected, but we believe that if the three particulars mentioned above are closely adhered to, either flame is reproducible within a maximum range of ± 20° from the mean.

The colour temperature of the Hefner flame is found to be 1910° K.

Although now somewhat out of fashion the Hefner lamp is still for many purposes a useful photometric standard and the cylindrical acetylene flame has been repeatedly recommended as a secondary standard(1). The spectral distribution of the light is however subject to doubt in both cases, as the following summary of published work shows.

The Hefner flame. Tumlirz(2) made a spectrophotometric comparison over the range 0·43–0·7 μ. between the Hefner radiation and sunlight, for which he assumed Langley's data. The resulting Hefner spectrum corresponds very roughly with a Planckian curve for 2100° K. Angström(3) found by bolometric measurements from 0·50 to 0·75 μ. that the colour temperature was 1830° K., and Leder(4) by spectrophotometric comparisons against black radiators at 1452° and 1616° K. confirmed Angström's result. Hyde and Forsythe(5) by comparison of tints on a Lummer-Brodhun head against standard filament lamps of known {voltage, colour-temperature} characteristics ascribe to the Hefner flame a colour temperature of 1880° K. Valentiner and Rössiger(6) by means of a monochromator and photoelectric cell made a spectrophotometric comparison of the brightest part of the Hefner flame and the electrode of a tungsten arc lamp, and concluded that the spectrum could be represented by a Wien equation for 1840° K. Dziobek and Hoffman(7) compared the Hefner flame with a black radiator both spectrophotometrically and by matching tints, and found its colour temperature to be 1900 ± 20° K. by the first method, and 1910 ± 5° K. by the second. The value most generally accepted appears to be 1875–1880° K.

The acetylene flame. According to Coblentz and Emerson(8) the flat (fishtail) acetylene flame is not sufficiently reproducible to be standardized, but the cylindrical (or rather cigar-shaped) flame gave "sufficiently concordant results to warrant its use as a standard in spectrophotometric work not requiring the highest precision". Guild(9) gives it a similar qualified approval. For photometric purposes Mees and Sheppard(1) and Jones(1) used a Bray burner consuming ¼ cubic foot of acetylene per hour with a flame 35 mm. high (Mees) or 50 mm. high and 3 mm. thick (Jones). Variation of the gas pressure from 8 to 10 cm. of water caused only 2·7-per-cent change in the photometric intensity of Jones's flame. Coblentz and Emerson examined radiometrically the spectra of the middle sections of two acetylene flames each consuming ¼ cubic foot per hour; the first, on a Crescent Aero tip, measured 35 mm. by 3 mm. under 7·5 cm. pressure; the second on a Bray tip was 50 mm. high at 9·0 cm. pressure. The spectra were identical throughout the visible region as far as 0·7 μ., beyond which wave-length the Bray flame began to show the higher intensity. Subsequent calculations(10) showed that the experimental values agreed within the limits of error ($\pm$ 1 to 2 per cent) with a Wien curve for 2360° K. except in the blue and violet beyond 0·48 μ., where they began to grow higher. Hyde, Forsythe and Cady(11) sent tungsten filament lamps, which had been standardized against black radiators, to the Eastman Kodak laboratory and to the Bureau of Standards, with a request that they should be colour-matched with an acetylene flame. The Kodak laboratory using a Bray tip at ¼ ft.^{3}/hr. found a matching voltage which gave a colour temperature of 2352° K.; the Bureau of Standards using the Aero tip found at first 2434°, and on re-test 2450° K. Hyde, Forsythe and Cady themselves made both colour-match and spectrophotometric comparison with their standardized lamps, and concluded that the Bray tip operating at 9 cm. pressure gave a flame of colour temperature 2360 ± 10° K. Lowering the gas pressure to 7·5 cm. changed this figure only a few degrees. Davis and Gibson(12) found that "a perfect colour-match between the acetylene light and the unfiltered electric light could not be obtained at any colour temperature". The nearest approach was somewhat over 2400° K. They point out that Coblentz's spectral energy values correspond, by the correlation method of Davis(13), with a colour temperature of 2413° K.; and by that of Priest by way of the spectral centroid, with 2388° K. The value 2360° recommended by Hyde, Forsythe and Cady is commonly accepted, but the above discrepancies clearly need an explanation. The confusion confronting inquiry into the literature of this subject is no doubt responsible for a misleading table on p. 401 of volume 19 of the *Handbuch der Physik* of Geiger and Scheel, which purports to give the spectral-energy ratios of the acetylene flame to black radiators at 2360° and 2450° K. The figures in this table were in fact calculated by Hyde, Forsythe and Cady to show that the uncorrected radiometric results first published by Coblentz did not correspond with the spectrum of a black body at any temperature.

During the past two years we have compared a cylindrical acetylene flame with a Hefner flame on a Hilger-Nutting spectrophotometer. In all seven sets of readings, each duplicated by two observers, have been taken. Corresponding sets of readings of the instrumental constants were taken, and it may be remarked that the flame is very suitable for this purpose on account of its axial symmetry. The procedure was that one observer made at least four optical settings at each wave-length, starting at 0·57 μ., progressing alternately towards the red and the violet by steps of 0·02 μ., and returning similarly by way of intermediate wave-lengths, so as to give a complete set of readings at 0·01-μ. intervals from 0·45 to 0·70 μ. The settings were read off by a second observer in order to preserve the sensitivity of the main observer's eye.

The acetylene was generated from commercial carbide (in accordance with the practice of Jones, Coblentz, Davis, etc.) and between 98 and 99 per cent of it was absorbed by bromine water. It was passed through towers containing respectively copper-sulphate crystals and anhydrous calcium chloride. Hydrides of phosphorus are always present in carbide-generated acetylene, but not in sufficient amount to affect the spectrum appreciably. No change could be detected in spectrograms from acetylene to which quantities of phosphine up to 1 per cent by volume had been added. The gas was stored overnight before use in a 5-ft.^{3} holder over water, and was burnt on a Bray Vika tip taking $\frac{1}{4}$ ft.^{3}/hr at a pressure of 9 cm. of water. The holder bell had compensating weights whereby the pressure was automatically kept within a range of ± 1 mm. The flame was from 49 to 50 mm. high and about 3 mm. thick in its central portion. A hood of the form described by Jones shielded from observation all except the central section from 15 to 25 mm. up the flame. The whole Hefner flame was observed in the other pupil of the spectrophotometer. The collimator slit was 0·5 mm. wide. The precision attainable with the Hilger-Nutting instrument has been considered in detail recently by Twyman and Lothian(14) and our experience leads us to think that they have certainly not over-estimated it. The average deviation of a single density reading from the arithmetic mean of its group was from 0·005 to 0·01 for wave-lengths from 0·48 to 0·66 μ., and increased to about 0·03 at 0·45 μ. Introduction of filters to stop stray light at the extreme wave-lengths made no detectable difference to the settings. The wave-length scale of the spectrometer was checked at a number of mercury and neon lines, and the density scale of the photometer was checked by means of a sector reading to 1 min. of arc.

The mean experimental values of the spectral ratios acetylene : Hefner, adjusted by a constant factor to unity at 0·57 μ., are given in column (2) of table 2, page 1038, and their logarithms are plotted in figure 1 alongside those calculated from the accepted colour temperatures 1875° and 2360° K. The slight inflections in the experimental curve at wave-lengths 0·49, 0·54–0·55 and 0·58–0·60 μ. are reproduced with great fidelity in all the tests. They do not appear to be an instrumental effect because, apart from the fact that they remain after allowing for the instrumental constant, they are not altered when the two flames are interchanged in the set-up.

The clear discrepancy between experimental and calculated ratios led us to compare each flame separately with the anode of a tungsten arc lamp which could be kept at a constant determinable temperature.

The arc was a pyrometric Pointolite having a bun-shaped anode about 8 mm.

in diameter, and the temperature of the outer surface of the anode was read on a Cambridge disappearing-filament pyrometer immediately before and after each comparison, the current being held constant within narrow limits by manual adjustment during the whole period. It was first necessary to see if the temperature

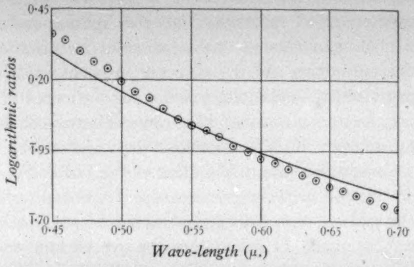


Figure 1. Logarithms of (acetylene : Hefner) ratios. ——— calculated for 1875° and 2360° K. ⊙ experimental comparison.

of the tungsten surface varied locally, and therefore an extension was fitted to the pyrometer tube so that different portions of an enlarged image could be observed. Table 1 gives uncorrected pyrometer readings taken (*a*) in the centre of the enlarged image; (*b*) about $\frac{2}{3}$ of a radius above the centre; (*c*) and (*d*) at similar distances along horizontal radii towards the sides nearest respectively to the lighting coil and the main cathode rod.

Table 1. Temperature-distribution over surface of tungsten anode

Position of arc on anode	Side	Centre
Current (A.)	0·95	1·00
Temperature (° K.) observed at (*a*) centre	1459	1477
(*b*) top	1458	1469
(*c*) coil side	1484	1472
(*d*) rod side	1442	1467
Weighted mean	1462	1474
Temperature observed on normal unenlarged image	1462	1479

The arc strikes at the side of the anode nearest the lighting-coil, and occasionally at small currents it fails to move over into the centre. The first column of figures shows that in this position there is a temperature-gradient from the coil side to the rod side. A vertical central strip of uniform temperature with hotter and cooler areas at its respective sides being assumed, a mean value, weighted as to area and brightness, was calculated and was found to agree with temperatures read from the small image when the extension tube of the pyrometer was not used. By temporarily increasing the current the arc could be made to travel to the centre of the anode, but there was a tendency to wander when the arc was run for long periods at a low current, and therefore the results at lower temperatures are considered less reliable than those at higher temperatures. When the arc was central on the anode the slight radial temperature-gradients were within the limits of reproducibility from day to day of the pyrometer readings. In actual working, therefore, the arc was always manoeuvred so as to settle in the centre of the back of the anode and the pyrometer was used without extension tube. A stop was fitted close to the bulb enclosing the arc, whereby all radiation from the lighting-coil and cathode rod (which project from behind the anode at the two sides) and from the lower edge of the anode where its supporting rod is connected, was shielded from the spectrometer. The pyrometer, which had been checked at the National Physical Laboratory, had a red glass filter, the transmission of which was determined. Its effective wavelength was 0·6475 μ. at 1910° K. The readings were corrected by means of Wien's equation for loss of brightness due to the glass of the bulb. For this purpose the bulb was broken at the end of the experiments and spectral transmission factors, given in column 6 of table 2, were determined on the area which had been exposed by the stop above mentioned. In calculating the overall loss, re-reflection at the surface of the anode was taken as 56 per cent from consideration of data by Coblentz(15) for cold tungsten and by Langmuir(16) for molten tungsten. The actual brightness temperature of the anode surface having been thus determined, the corresponding colour temperature was calculated from the tables of Forsythe and Worthing(17). Forsythe(18) has shown that the emission of tungsten is very nearly that of a black body at the same colour temperature, and we therefore assume, as the basis of our comparisons, energy curves calculated by Wien's equation for the determined colour temperatures.

The comparisons of the two flames against the tungsten arc were made in two ways, (*a*) photographically with spectrometer alone, and (*b*) visually with spectrophotometer.

(*a*) With exposures suitable for the blue end of the spectrum, the green and yellow are so over-exposed that the spectra can only be compared over a small range. We therefore inserted in the camera just in front of the plate a filter covering the spectral region from about 0·54 μ. onward, whereby the intensity of that part was reduced to about the same photographic intensity as that of the blue end. Suitable filters were, for the acetylene flame, Wratten No. 78AA; and for the Hefner flame, Wratten No. 37 in double thickness. The method of making the comparison was then the following. A series of equal exposures of the flame were made, so that the spectrograms were spaced at equal intervals from top to bottom of the plate. This was easily done by means of a graduated rack motion fitted to the camera. The flame was then replaced by the tungsten arc running at some selected current. A rotating sector was interposed whereby the illumination was reduced to an intensity rather greater than that of the flame in the previous exposures. A series of spectrograms of the arc was then taken in the intervals left on the plate between the flame spectrograms, and at each exposure the intensity of the illumination reaching the spectrometer was reduced in 5-per-cent steps by means of the sector. Under these conditions, if the arc happened to have the same colour temperature as the flame the first pair of spectrograms would show the arc spectrum uniformly deeper than the flame spectrum; the last pair would show it uniformly lighter than the flame spectrum; and there would be some intermediate pair which would match each other exactly throughout the spectrum. If, however, the arc had a colour temperature differing from that of the flame, one end of its spectrum would be reduced to equality with the flame spectrum sooner than the other, which would require one or more further stepwise reductions of the sector aperture before it matched. The number of steps between the matching of the two ends of the spectrum is thus a measure of the difference in colour temperature between arc and flame. Such sets of paired spectrograms were taken with the arc held at a series of temperatures, and for each temperature the number of steps measuring the difference between the matching pair for the red and that for the blue end was noted. The differences were plotted against arc temperatures, figure 2, and the point at which the best-fitting line cut the temperature axis was taken as that at which the two colour temperatures were identical.

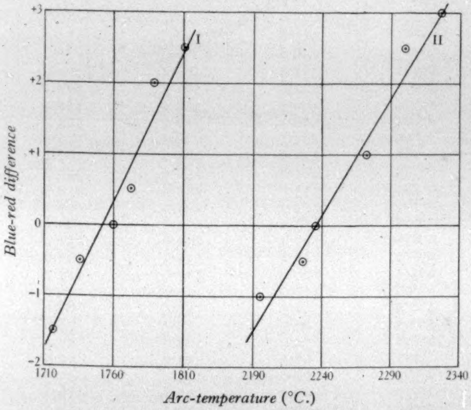


Figure 2. Curve I, Hefner; curve II, acetylene.

In this way the Hefner flame was found to match the arc running at a pyrometer reading of 1752° K. This, corrected for the effect of the glass bulb, gives the brightness temperature 1758° K. which, from Forsythe's tables, corresponds with a colour temperature of 1913° K. The acetylene flame was found similarly to match a colour temperature of 2519° K.

(*b*) The above photographic method takes no account of any possible selective transmission by the glass of the arc bulb. Such selectivity is very slight, as may be seen from the transmission factors in table 2, column 6, and its effect is allowed for in the following visual comparisons of each flame against the arc, made on a Hilger-Nutting spectrophotometer. The settings were made as already described by each of two observers, the other meanwhile controlling the arc current and taking down the instrumental readings. The arc was run at a current corresponding with a colour temperature of 2520° for the acetylene flame, and 1925° K. for the Hefner flame. The spectral factors averaged from four sets of readings by each observer were corrected in respect of the transmission factors of the bulb glass and then applied to spectral-energy values calculated from Wien's equation for the appropriate temperature. The results giving the spectral distribution of the radiant energy from each flame are tabulated in columns 3 and 5 of table 2, while column 4 shows the Hefner spectrum calculated by means of factors obtained from direct comparison of the two flames, on the assumption that the acetylene flame radiates as a black body at 2520° K. At the foot of each column is the black-body temperature having the same centroid over the range 0·45–0·69 μ. as the tabulated spectrum. In figure 3 the experimental values from columns 3 and 4 have been plotted alongside Wien curves for 2520° and 1910° K. respectively. The figures in column 5

Table 2

1 Wave-length (μ.)	2 Ratio of acetylene to Hefner, by direct comparison	3 Acetylene flame	Hefner flame by comparison with		6 Trans-mission factors of bulb glass
			4 Acetylene flame	5 Tungsten arc	
0·72	—	257·6	—	—	—
0·71	—	246·0	—	—	0·914
0·70	0·561	234·4	407·4	—	0·916
0·69	0·573	221·6	381·5	384·1	0·918
0·68	0·594	209·4	349·9	352·4	0·918
0·67	0·621	198·2	318·4	323·6	0·920
0·66	0·649	187·1	289·1	294·4	0·920
0·65	0·668	177·0	265·2	267·9	0·923
0·64	0·705	165·8	236·9	240·7	0·923
0·63	0·734	155·2	213·8	214·3	0·923
0·62	0·769	145·7	191·2	190·7	0·923
0·61	0·807	137·1	169·8	169·4	0·923
0·60	0·836	127·2	152·6	151·2	0·923
0·59	0·879	117·2	134·3	133·0	0·923
0·58	0·925	108·2	117·6	115·7	0·923
0·57	1·000	100·0	100·0	100·0	0·923
0·56	1·055	91·3	86·5	86·2	0·923
0·55	1·105	82·7	75·3	74·6	0·920
0·54	1·156	74·9	64·9	63·2	0·918
0·53	1·253	67·3	54·0	52·8	0·916
0·52	1·365	60·0	44·4	43·9	0·916
0·51	1·439	53·2	37·4	36·3	0·914
0·50	1·560	46·9	30·5	29·9	0·914
0·49	1·742	41·3	23·9	24·4	0·916
0·48	1·826	36·2	19·9	19·7	0·918
0·47	2·004	31·3	15·6	15·7	0·918
0·46	2·183	26·9	12·3	12·5	0·918
0·45	2·307	22·9	9·9	9·8	0·916
0·44	—	19·4	—	—	—
0·43	—	16·2	—	—	—
Colour temperature (° K.)		2520	1912	1897	—

lead to a slightly lower colour temperature than those in column 4, and they are considered to be less reliable on account of the tendency of the arc to wander occasionally when run at low currents.

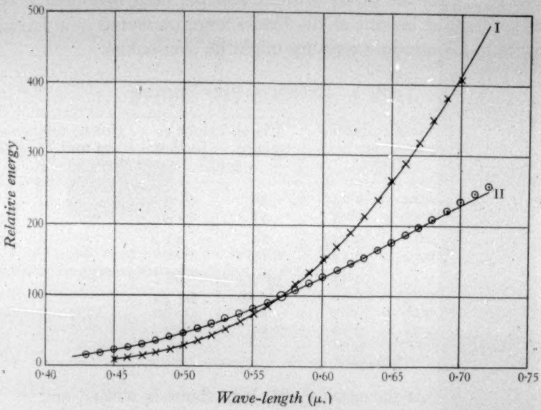


Figure 3. Curve I, black body at 1910° K.; × Hefner flame. Curve II, black body at 2520° K.; ⊙ acetylene flame.

The values given in table 3 for the colour temperature of the entire Hefner flame are higher than the accepted value 1875–1880° K.; they agree with the estimates of Dziobek and Hoffmann referred to above.

Table 3. Summary of colour temperature determinations

Method	Hefner (° K.)	Acetylene (° K.)
Photographic comparison against tungsten arc	1913	2519
Spectrophotometric comparison against tungsten arc	1897	2520
Spectrophotometric comparison against acetylene flame	1912	—

The colour temperature of the central section of our acetylene flame is much above any other published value, and we sought an explanation of the discrepancy by examining the flames produced by various burners. The burners are manufactured by Geo. Bray and Co. Ltd., of Leeds, in two types under the trade names Ota and Vika, and in sizes consuming $\frac{1}{8}$, $\frac{1}{4}$ and $\frac{1}{2}$ cubic feet per hour. The makers state that the consumption figures stamped on the sockets show the hourly consumption of gas at a pressure of 3 in. of water. Previous workers have used pressures of 75 and 90 mm. of water, and we therefore tested the consumptions of several burner tips at these pressures and compared some of the flames spectrophotometrically with that used in the tests described above, using always the central portions of the flames at heights of 1·5–2·5, 2·5–3·5 and 1·0–1·5 cm. respectively above the bases of the Vika $\frac{1}{4}$-ft^3, Ota $\frac{1}{4}$-ft^3 and Vika $\frac{1}{8}$-ft^3 flames. Table 4 contains the results. The consumptions were measured by means of a water-sealed meter passing $\frac{1}{12}$-ft^3 per revolution and calibrated with the $\frac{1}{12}$-ft^3 bottle as prescribed by the Gas Referees. They are given in cubic feet per hour measured at 60° F. and 30 in. of mercury. The heights of the flames were measured in a darkened room since otherwise the faintly luminous tip might be overlooked.

Table 4. Details of Bray burners

Maker's description		Consumption at (ft^3/hr.)		Flame height at (cm.)		Colour temperature at (° K.)	
		75 mm.	90 mm.	75 mm.	90 mm.	75 mm.	90 mm.
Vika $\frac{1}{4}$-ft^3	(*a*)	0·22	0·245	40	48–49	2535	—
	(*b*)	0·22	0·25	41	49–50	—	(2520)
Vika $\frac{1}{8}$-ft^3	(*a*)	0·11	0·127	18	22	—	2670
	(*b*)	0·11	0·126	19	21–22	—	—
Ota $\frac{1}{4}$-ft^3	(*a*)	0·27	0·30	52–53	60	—	—
		0·25	0·28	48–49	54–55	—	—
	(*b*)	0·26	0·29	52	59–60	2377	2407
	(*c*)	0·26	0·30	52	59	—	—
	(*f*)	—	0·29	—	58	—	2377

The results show that the small ($\frac{1}{8}$-ft^3) Vika flame is whiter, and the Ota $\frac{1}{4}$-ft^3 flame is yellower, than the $\frac{1}{4}$-ft^3 Vika flame. The differences are clearly attributable to differences in primary aeration. An acetylene burner must admit enough primary air to stiffen the flame and prevent carbon-deposition. It is well known that as the primary aeration is increased the flame-temperature rises, so that the flame grows continuously whiter although its total luminosity soon reaches a maximum and begins to fall. Thomas(19) has shown that, other things being equal, the volume of air entrained per unit volume of gas delivered increases as the size of the jet diminishes; and the colour temperatures found are in fact in reverse order of the consumptions, thus: Vika $\frac{1}{8}$ > Vika $\frac{1}{4}$ > Ota. Thomas also showed that air-entrainment increases with delivery pressure, and this is exemplified in the case of Ota burner (*b*), the colour temperature of which is raised by increasing the pressure from 75 to 90 mm.

The majority of previous workers used burners consuming $\frac{1}{4}$-ft^3/hr. at a pressure of 90 mm., with a flame 50 mm. high. The Vika tip which we have used as our standard conforms nearly to this specification. If these three easily measured magnitudes—gas-pressure, consumption and height of flame—are fixed, it appears that the flame is sufficiently reproducible for use as a spectral standard when the highest accuracy is not required. It is however necessary to select the tips carefully. For example, the Vika $\frac{1}{4}$-ft^3 tip (*a*) in table 4 is slightly under size, and has a slightly but distinctly higher colour temperature than the standard (*b*). Some tips are not quite truly set in the socket, and give flames which are not vertical. The dimensions of the primary airholes are not very critical, at any rate within certain limits. Our standard Vika tip has four holes each of diameter 1·2 mm. We selected two tips having the same consumption and flame height, closed up one of the four holes of

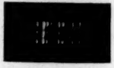


Figure 4. Unenlarged synchronous spectrogram of moving striations in neon in orange and red regions.

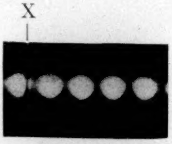


Figure 5. Moving striations in neon.

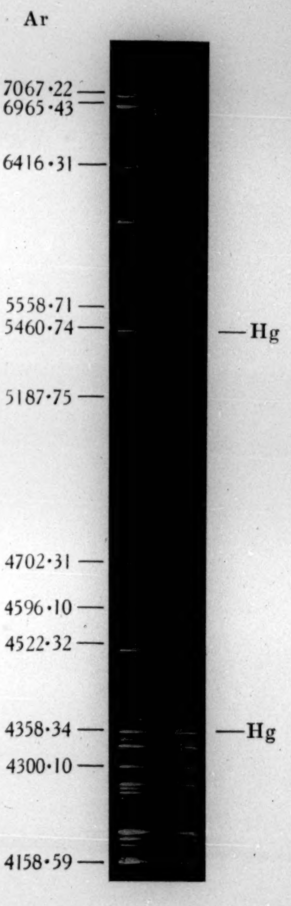


Figure 7. Synchronous spectrogram of striations in argon-mercury mixture.

the
met
252
252
wit
dri

He
tha
tab
to
of
fal
ov
hi
fla
o
v

(
a

the first, and enlarged those of the other to 1·3 mm. diameter. On spectrophotometric comparison with the standard the former gave a colour temperature of 2522° K., which was indistinguishable from that of the standard, and the latter 2527° K. Variations so great as these intentional changes of size would not be met with in practice. The airholes can in any case be easily gauged by means of standard drills.

The Ota tips approach their rated consumption only at the lower pressure. Hence their flame, as has already been stated, is larger, less aerated and cooler than that of the Vika tip. Their airholes are also slightly smaller. The Ota tip (*a*) in the table possessed two definite consumption-rates. This phenomenon is possibly due to some irregularity in the bore causing instability of flow, but as the temperature of the gas in the bore is unknown it is impossible to say definitely that its velocity falls within the critical range. The tip (*f*) was kindly lent by Dr Forsythe from his own apparatus. Our value for the colour temperature of this flame is slightly higher than that given by Dr Forsythe, which was 2360° K. Ota tip (*b*) gave a flame of 2407° K. when tested over a length of 1 cm. With slot reduced to expose only about 4 mm. the result was 2402° K. The difference is within the limits of variation of individual tests.

ACKNOWLEDGMENTS

This work was done in the Somerville Laboratory of the South Metropolitan Gas Company, and we are indebted to the President, Dr Charles Carpenter, C.B.E., and the Board of Directors for permission to publish it.

REFERENCES

(1) Mees and Sheppard. *Photogr. J.* **50**, 287 (1910).
Jones. *Trans. Illum. Engng Soc.*, N.Y., **9**, 716 (1914).
(2) Tumlirz. *S.B. Akad. Wiss. Wien*, **98**, Abt. 2*a*, 1121 (1889).
(3) Angström. *Phys. Z.* **5**, 457 (1904).
(4) Leder. *Ann. Phys.*, Lpz., **24**, 305 (1907).
(5) Hyde and Forsythe. *Trans. Illum. Engng Soc.*, N.Y., **16**, 419 (1921).
(6) Valentiner and Rössiger. *Ann. Phys.*, Lpz., **76**, 785 (1925).
(7) Dziobek and Hoffman. *Z. InstrumKde*, **47**, 327 (1927).
(8) Coblentz and Emerson. *Bull. U.S. Bur. Stand.* **13**, 355 (1916–17).
(9) Guild. *Proc. Opt. Conv.* Part 1, pp. 75–6 (1926).
(10) Coblentz and Emerson. *Bull. U.S. Bur. Stand.* **15**, 639 (1919–20).
(11) Hyde, Forsythe and Cady. *Phys. Rev.* **14**, 379 (1919).
(12) Davis and Gibson. *Misc. Publ. U.S. Bur. Stand.* No. 114, p. 26 (1931).
(13) Davis. *Bur. Stand. J. Res.*, Wash., **7**, 689 (1931).
(14) Twyman and Lothian. *Proc. phys. Soc.* **45**, 643 (1933).
(15) Coblentz. *Bull. U.S. Bur. Stand.* **7**, 197 (1911).
(16) Langmuir. *Phys. Rev.* **6**, 138 (1915).
(17) Forsythe and Worthing. *Astrophys. J.* **61**, 146 (1925).
(18) Forsythe. *J. opt. Soc. Amer.* **7**, 1115 (1923).
(19) Thomas. *Phil. Mag.* **47**, 1048 (1924).

621.385.334

THE ELECTRON-OSCILLATION CHARACTERISTICS OF AN EXPERIMENTAL PLANE-ELECTRODE TRIODE

By R. A. CHIPMAN, B.Sc., M.Eng.,
Cavendish Laboratory, Cambridge

Communicated by Prof. E. V. Appleton, F.R.S., June 1, 1935. *Read October* 4, 1935

ABSTRACT. To investigate the characteristics of electron oscillations generated under simplified conditions, an experimental plane-electrode triode has been constructed. Its plane emitting surface is 25 mm. in diameter. The cathode-plate distance can be continuously varied while the valve is in operation. External circuits may be connected to the electrode leads in such a way that the resultant oscillating circuit is effectively continuous from the electrodes to the terminating condenser. With no external circuit connected to the valve, oscillations are maintained whose wave-length varies with grid voltage exactly according to the Barkhausen equation $\lambda^2 V = \text{constant}$. Their mechanism appears to be confined to the grid-cathode space. They are produced only when the grid current is space-charge-limited. When an external circuit is used, oscillations are maintained at very low emissions. Their wave-length is always identical with some resonant wave-length of the circuit. Their mechanism may be confined to the grid-plate space, or may extend over the cathode-plate distance. First and second-order dwarf waves are observed. It is shown that much more complex experimental results can be analysed on the basis of the conclusions drawn from the simplified experiments. A brief discussion of the applicability of existing theories to the simplified experiments is given.

§ 1. INTRODUCTION

In recent years a very large amount of experimental work[1] has been carried out in the investigation of the mechanism and characteristics of electron oscillations. A satisfactory solution of the problems involved has not, however, been found. The phenomena observed have been so complex and so diverse that it has not been possible to penetrate through them to common fundamental laws. The many theories which have been advanced are based on a wide variety of mechanisms and, while each has had some degree of experimental justification, there have been numerous instances in which investigators have been unable to duplicate the results upon which others have founded such theories.

Consideration suggests that this confused situation can only be due to the facts that, with a few minor exceptions, the experiments have been made entirely with commercial valves, and that different experimenters have used different types of valve. Valve-construction parameters play a very critical part in the generation of electron oscillations. Potapenko[2] and others have shown that even valves superficially identical may have quite different oscillation characteristics. Further, the structure of the electrodes and the electrode leads in commercial valves is not amenable to simple theoretical treatment. Hence the results that have been obtained cannot in general be compared directly with one another or with theoretical results based on simplifying assumptions.

In the present investigation an attempt has been made to reduce the problem experimentally to its simplest terms. A plane-electrode experimental valve has been designed to approximate as closely as possible to the theoretically simplest arrangement, and to eliminate indeterminate geometrical parameters. The aim has been to see whether the phenomena observed under these idealized conditions are explainable by any simple theory.

In this paper there are described only the basic results that have been obtained, with the number of critical variables reduced to an absolute minimum. Much more complex effects are observed when other variables are allowed to become important. Operating values for these minor variables have been so chosen as to make their influence either negligible or easily calculable. The results described have all been reproduced several times in essentially the same form, if not in exact detail, and it is claimed that they can be repeated by any experimenter.

§ 2. THE EXPERIMENTAL APPARATUS

The details of the experimental valve used are given in figure 1 (*a*). The Pyrex envelope *P* is sealed by white wax into a circular slot a few millimetres deep in the brass ring *B*. A circle of copper tubing soldered into another slot in the ring permits water cooling of the wax seal. The square of plate glass *D*, through which the leads to the electrodes pass, is sealed with tap grease to the polished face of the ring *B*. Plasticene around the circumference of the ring completes the seal.

The electrode leads are nickel rods 3 mm. in diameter. They are threaded for 2 in. of their length at the ends nearest the electrodes, and are fastened to the glass *D* by a washer and nut on either side. White wax is used as seal. Two leads are required for the cathode-heating coils and one each for the three electrodes. The grid lead has two adjustable joints, so that the grid-cathode distance can easily be varied on dismounting the valve.

The plane anode (which will be referred to henceforth simply as the plate, since it is not in this work a positive electrode) is fixed to a short length of Pyrex tubing which slides vertically in the tube *T*. A silk-thread suspension allows the vertical height of the plate to be varied by rotation of the rod *R* through the ground joint. Motion through a distance of 2–3 cm. is possible. The plate lead consists of a short length of nickel tube 3 mm. in inside diameter, pivoted at the plate, and telescoping on to a piece of 3-mm. rod pivoted just inside the glass *D*. With this construction the glass *D* carrying the grid and cathode can be dismounted, leaving the plate in position. Inter-electrode distances are measured with a microscope sliding on a vertical rod outside the envelope.

The constructional details of the indirectly-heated plane cathode are shown in figure 1 (*b*). The component parts are all mounted on a small soapstone base which provides the insulation between the heating supply and the cathode surface. There are two heating-coils, each requiring about 15 cm. of tungsten wire 0·5 mm. in diameter, wound into 20 turns 2·5 mm. in mean diameter. The box containing the heaters is circular, 25 mm. in diameter, and formed from nickel sheet of thickness 0·5 mm. Similar sheet is used for the two radiation shields. The upper edge of the cathode box is spread to form a flange, and a circle of 0·1-mm. nickel foil is spot-welded to this flange to form the emitting surface. A suspension in water of finely

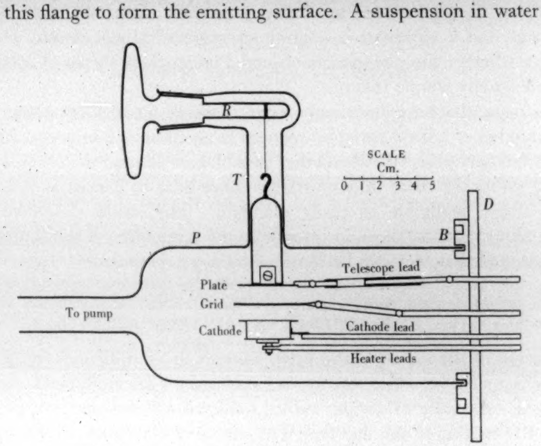


Figure 1 (*a*). The experimental plane-electrode valve.

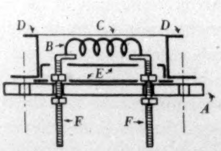


Figure 1 (*b*). Constructional details of the plane cathode. *A*, soapstone base; *B*, tungsten heating coils; *C*, emitting surface; *D*, non-emitting guard ring; *E*, radiation shields; *F*, current leads.

ground barium hydroxide painted on to the foil, and dried in air, gives emission very easily and quickly, provided that the surface is clean and the system well sealed. The 5-mm. flange around the actual emitting surface serves as a guard ring to eliminate any possible distortions of the electrostatic field that might be caused by the electrode leads, or other edge effects.

The grid used throughout the experiments described in this paper consists of parallel 0·5-mm. nickel wires spaced at 1·5 mm. between centres, spot-welded at either end to a 1·5-mm. nickel wire frame. The frame is 4 cm. square. Grid-transparency (67 per cent in this case) has been found experimentally to be a very important parameter, and the results described here are not in every case characteristic over a wide range of this variable. The plate used throughout consists of a circle of sheet nickel 3·5 cm. in diameter.

The parallel-wire external circuits which can be connected to the valve are all constructed of brass telescope tube, of intermediate diameter 7 mm. Connexion to the electrodes is made by threading the smaller brass tube over a short length of brass rod which in turn threads over the nickel electrode leads. The spacing between the two conductors is always the same as that between the electrode leads where they pass through the glass plate *D*, whichever pair of electrodes is being used. The terminating condensers are Sangamo 0·0005-μF., and are fastened to very short lengths of brass strap soldered to the ends of the brass tubes.

With this arrangement the resonant circuit in which oscillations may occur consists effectively of parallel rods extending continuously from the electrodes to the terminating condenser. It can thus be accurately analysed and the indeterminacy introduced by the electrode-seal in commercial valves is eliminated. The change in circuit constants just outside the glass plate is not serious, and no indication of spurious reflection from this point has ever been obtained.

The leads carrying the heating-current to the valve from the transformer are connected by a 0·001-μF. condenser immediately outside the plate *D*, and by two other similar condensers spaced at 15 cm. along the length of the leads away from the valve. This effectively prevents the leads from acting as a resonant circuit at any of the frequencies generated experimentally. Radiofrequency chokes, used wherever necessary, are space-wound of 32-gauge wire on 5-mm. glass tubing, and fulfil their purpose satisfactorily.

By detecting oscillations generated by this valve with a heterodyne receiver, it has been shown that the magnetic field produced by the cathode-heating coils has no observable influence on the nature of the oscillations. When the heating current is turned off the cathode remains hot enough to emit for several seconds, and the detected oscillations remain unchanged in frequency and intensity for at least 1 or 2 sec., before they gradually decay away. Nickel loses its ferromagnetic properties at a temperature far below that at which the cathode normally operates, so that no effects can be attributed to magnetization of the cathode box.

For measuring wave-length, absorption wave-meters similar to those described by Moore(3) have been calibrated down to 60 cm. Where these cannot be used, a crystal circuit of the type described by Chapman(4) is very loosely coupled to the valve. This circuit may also be used to give an indication of oscillation-intensity. The coupling is so loose that resonance in the crystal circuit does not visibly affect the plate current of the oscillator. There can be no question of appreciable interaction between the two circuits.

§ 3. THE EFFECTS OF GAS PRESSURE

Nettleton in 1922 published a paper(5) describing "electron" oscillations which disappeared when the pressure in the valve was reduced to $5 \cdot 10^{-5}$ mm. of mercury. This immediately suggested that the oscillations were not produced solely by electrons, but required the presence of positive ions. Tonks and Langmuir(6) have shown experimentally that an ionized gas can produce oscillations of the correct order of frequency. The pressure required to give a sufficiently high ionic density must be about 10^{-3} mm. It is possible, therefore, that at such pressures the positive ions do play an essential part in the generation of oscillations in a positive-grid triode.

For the usual cases of oscillations in positive-grid circuits, however, it has been proved very conclusively by Grechowa(7) and by Rindfleisch(8) that the principal effect of gas is to damp the oscillations. These workers found, in independent experiments, that the oscillation-intensity increased asymptotically to a maximum as the pressure decreased down to 10^{-8} mm.

Before experiments were started with the present plane-electrode valve a pressure test similar to those described in the above papers was carried out on an AT 40 valve. The object was to determine an upper limit for the pressure at which oscillation-intensity and wave-length can be regarded as effectively independent of the presence of the gas. The AT 40 is an ordinary commercial valve of cylindrically symmetrical construction, having an anode-diameter of 12·8 mm. and a grid-diameter of 5 mm. A glass tube was sealed into the valve at the tip, a small bulb containing charcoal was joined to this tube, and the whole was sealed on to a diffusion pump. After baking of the glass, degassing of the electrodes by electron bombardment, and heating of the charcoal for some time, the valve was calibrated as a positive-grid ionization gauge against a McLeod gauge for several values of pressure down to 10^{-5} mm., the lowest pressure readable with any accuracy on the McLeod gauge. After further heating and pumping the valve was sealed off and removed from the pump. Linear extrapolation of the calibration curve of an ionization gauge being assumed to be valid, it was then found that liquid air applied to the charcoal would reduce the pressure in the valve to slightly below $2 \cdot 10^{-7}$ mm. By surrounding the charcoal bulb with ice, boiling water, or a heated oil bath, other values of pressure could be maintained steadily. Values of optimum grid voltage, wave-length, and oscillation-intensity were taken at each pressure, the external circuit remaining fixed as a three-quarter-wave system. The results are shown in figure 2.

These results are in general agreement with those previously mentioned. The oscillations disappear at a pressure of about $5 \cdot 10^{-4}$ mm., owing probably to the electrons losing too much energy to the gas molecules. The considerable increase of wave-length with pressure may be due to a combination of two effects. The inter-electrode capacity, which forms a part of the resonant circuit, varies with the ionic density between the electrodes, so that the resonant frequency of this circuit is being changed. The presence of the ions and molecules must also affect the electron transit time. The fact that the experimental grid-voltage curve falls much more rapidly than the theoretical grid-voltage curve corresponding to the experimental wave-length curve suggests that the effect of the gas is to reduce the transit time.

For these fairly intense oscillations it is seen that pressures up to 10^{-4} mm.

do not affect the intensity appreciably, and affect the wave-length by only about 2 per cent. Oscillations of very small intensity, however, generated at low emission current, disappeared when the pressure was increased to about $5\cdot10^{-5}$ mm., and their wave-length at this pressure was about 4 per cent higher than at the minimum pressure. They remained unaffected for pressures up to $3\cdot10^{-5}$ mm. It may therefore reasonably be considered that at pressures below 10^{-5} mm. both oscillation-intensity and wave-length may be regarded as independent of the gas, whatever the actual magnitude of the oscillation intensity.

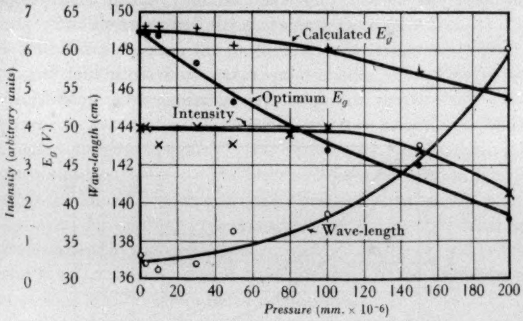


Figure 2. Effects of gas pressure on the wave-length and intensity of electron oscillations generated in a cylindrical-electrode triode. Filament current and external circuit-length constant.

Because of the numerous wax and glass seals in the plane-electrode valve, the size of the electrodes, and the fact that the electrodes are exposed to the atmosphere each time the valve is dismounted, the attainment of a pressure very much below 10^{-5} mm. would require that a great deal of time be spent on checking seals and degassing the electrodes. The pump used is a steel mercury-vapour condensation pump, and the connexion from the pump to the valve consists of 20-mm. Pyrex tubing. This equipment with a liquid-air trap will maintain the pressure at about 10^{-5} mm. even when there are small leaks present, and will produce a much lower pressure if the system is perfectly sealed. It will be considered throughout, therefore, that gas pressure is not influencing the results to any measurable extent.

§ 4. BASIC ELECTRON OSCILLATIONS

The most fundamental generator for the production of true electron oscillations should be one which does not depend on any resonant inductance-capacity circuits. Oscillations obtained from such a generator must be regarded as being characteristic of some property of the electron motion alone, and not simply as oscillations inevitably produced in a resonant circuit of zero effective resistance.

Barkhausen and Kurz visualized oscillations of this type in their original paper on the subject⁽⁹⁾, although they did not actually observe them experimentally.

They suggested that the period of the detected oscillations might be the same as the period of vibration of the electrons through the grid of the valve, and showed that this would result in a relation between wave-length and grid voltage given by $\lambda^2 V=$ constant. Most commercial valves will not produce oscillations whose wave length follows this equation when the grid voltage is varied over a wide range, probably because of resonant circuits inherent in the electrode structure.

To look for such oscillations with the present plane-electrode valve, the external circuit was removed and the potentials were applied through radiofrequency chokes connected to the electrode leads immediately outside the envelope. This left 12 cm. of nickel rod with no terminating condenser as the only possible high-frequency circuit external to the electrodes. Its maximum resonant wave-length would be 24 cm. With the cathode-temperature sufficiently high to provide space-charge-limited grid-current throughout, oscillations were maintained with this circuit for all grid-voltages between 35 volts and the highest value tested, corresponding to a wave-length range of from 220 to 130 cm. The oscillation intensity, measured on a loosely coupled crystal circuit, increased continuously with the energy-input to the valve. The curve of wave-length against grid voltage is shown in figure 3 and evidently lies close to a curve obeying the Barkhausen relation.

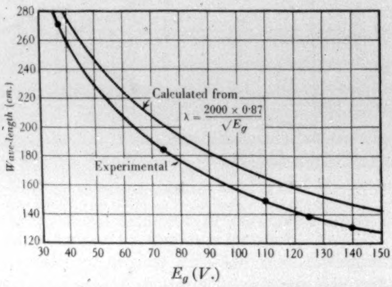


Figure 3. Variation of wave-length with grid voltage for oscillations generated by the plane-electrode valve with no external circuit. The cathode-temperature is high enough to supply space-charge-limited grid current over the whole range of E_g. Cathode-grid distance, 8·7 mm.; grid-plate distance, 10·4 mm.

The agreement between these two curves means that the period of the oscillations is linearly related to some electron transit time over the whole range of the experimental curve. It is found that the period is almost exactly equal to the transit time from cathode to grid and back, space-charge being neglected in the calculation. The oscillation mechanism seems to be confined to the space between cathode and grid, and three further facts confirm this view. Firstly, for three different values of grid-cathode distance the period of the oscillations agreed consistently with the return transit time across this space. Secondly, variation of the grid-plate distance from 5 to 17 mm. in a typical case does not change the wave-length by a detectable amount, although it affects the intensity considerably. Finally, earthing either grid or cathode leads just outside the valve stops the oscillations, while earthing the plate lead has only a very small effect on their intensity.

Reducing the cathode temperature at any fixed grid voltage causes the oscillations to disappear when the grid current falls appreciably below the space-charge-limited value. Their intensity remains constant at its maximum value for all cathode-temperatures above that required to produce this limited grid current.

Applying a positive plate-potential reduces the intensity of the oscillations, and they disappear when the plate is sufficiently positive to collect most of the electrons passing the grid. Thus although the position of the plate does not affect the fundamental mechanism of the oscillations, its presence is necessary to provide a returning stream of electrons in the grid-cathode space.

The theoretical curve of figure 3 is based on an electron transit time derived by neglecting space-charge. Megaw(10) has shown that calculations for a plane-electrode valve similar to those of McPetrie(11) for a cylindrical valve indicate an increase in transit time of just 50 per cent when complete space-charge is present. There is, however, no existing experimental proof of this fact for a positive-grid triode. The agreement of the two curves of figure 3 may prove either that the effect does not exist at all, or merely that it does not apply to the particular vibration in the electron motion which causes these oscillations.

A few commercial cylindrical electrode valves are found to give results similar to those shown in figure 3, with a slight change in experimental procedure. Filament-temperature is found to be a critical variable with these valves, and for each grid voltage there is an optimum value of heating-current. This optimum value is in general that which places the operating point on the upper knee of the grid-current saturation curve. Figure 4 shows a curve of wave-length against grid voltage obtained from an Osram R valve, together with a curve obeying the Barkhausen relation. This valve has a plate-diameter of 10 mm., a grid-diameter of 4·5 mm., and a grid-transparency of about 80 per cent. Circuit conditions are identical with those described for figure 3, chokes being connected in all the leads immediately outside the valve envelope. Heating-current has been adjusted to the optimum value for each experimental point.

Here again it is evident that the transit time of the electrons is the factor controlling the wave-length. Calculation shows that the transit time corresponding to the oscillation period in this case is that from filament to plate and return. This calculation is confirmed by the fact that earthing either plate or cathode terminals stops the oscillations.

Figures 3 and 4 may both be obtained equally well if the chokes are removed from the potential leads, provided these leads are kept well separated, so that no resonant circuit of low damping is formed by them.

The conclusion to be drawn from the results of this section is that the source of the basic oscillations lies in the space charge itself, and that they are entirely independent of and uncontrolled by any external resonant circuits. They are in fact the oscillations which have been referred to throughout the literature as Barkhausen-Kurz oscillations in contrast to Gill-Morrell oscillations, and have been sought by a number of experimenters, usually with negative results. Moore(13) and Kapzov and Gwosdower(12) have shown some evidence for their existence over narrow ranges of grid potential.

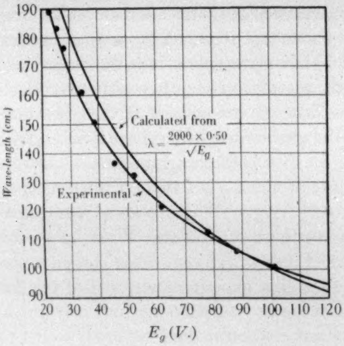


Figure 4. Variation of wave-length with grid voltage for oscillations generated by a cylindrical-electrode valve with no external circuit. The filament current is adjusted to an optimum value for each grid voltage. Osram R valve. Grid-radius, 2·2 mm. Plate-radius, 5·0 mm.

The production of oscillations by a space-charge-limited current has been discussed theoretically by Benham(14) and Llewellyn(15), but the present results cannot be regarded essentially as a confirmation of their theories. Benham's theory predicts oscillations in a diode, while oscillations could not be obtained from the plane-electrode valve unless electrons were allowed to return through the grid to the grid-plate space. Llewellyn's calculations are applied to a positive-grid triode, but the complex and varying relation between oscillation-period and electron transit time given by his equations cannot be compared with the simple and exact ratio found in these experiments.

§ 5. OSCILLATIONS WITH EXTERNAL CIRCUITS

When a low-resistance parallel-wire external circuit having suitable resonant wave-lengths is connected to any pair of electrodes of the plane-electrode valve, the experimental results assume several new aspects. It is found that the wave-length of all oscillations obtained lies very close to one of the resonant wave-lengths of the circuit, and variation of any parameter except circuit-length causes only very small continuous changes of wave-length. Since the electrodes are always approximately at a voltage antinode of the standing waves on the circuit, the resonant wave-lengths of a circuit of length L terminated by a low impedance condenser are

$4L$, $4L/3$, $4L/5$, etc. It is also found that space-charge-limited grid current is no longer necessary for the production of oscillations. Grid-currents as low as 1 per cent of the limited value may be sufficient.

A typical experimental curve is that shown in figure 5. This is taken with a circuit of fixed length connected to the grid and plate terminals. Grid voltage, grid-current and cathode-grid distance are maintained constant. The curve shows variation of wave-length and plate current (as an approximate measure of oscillation-intensity) with grid-plate distance. The grid current is only 1·5 per cent of the saturation value for the voltage and dimensions involved, so there can be no possible effects of space charge on the electron-motion.

There are seven distinct regions of oscillation in the curve, across each of which the wave-length is constant within 1 cm. The various wave-lengths correspond to the 1st, 2nd, 3rd, 4th and 5th harmonics of the circuit-length—in other words, the standing waves comprise 3, 5, 7, 9 and 11 quarter wave-lengths respectively.

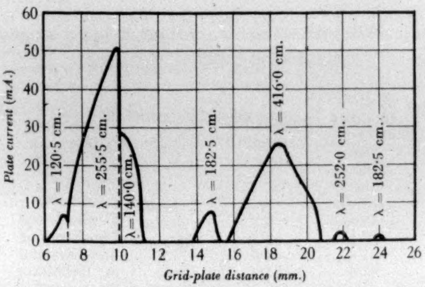


Figure 5. Oscillation regions obtained with the plane-electrode valve connected to an external circuit, on varying the grid-plate distance. All other variables are held constant throughout. External circuit-length from terminating condenser to centre of cathode, 310 cm.; grid potential, 144·2 V.; grid current, 0·75 mA.; cathode-grid distance, 5·5 mm.

The quantity $\lambda\sqrt{E_g}/2000$ expresses, for any oscillation generated in the absence of space charge, the distance across which the return electron transit time for the grid voltage E_g is equal to the period of the oscillations of wave-length λ. It may thus be taken as indicating which pair of electrodes is involved in the generation of any particular region. The value of this product for the seven regions of figure 5 together with the most probable analysis of the electrodes involved is given in table 1.

The analysis shows that the two most important oscillation regions appear to involve a cathode-plate electron-transit, while the three regions second in importance fit a grid-plate electron transit, and the weakest oscillations are those which seem to be related harmonically to the electron-motion. The conception of oscillations whose period is a simple fraction of the electron transit time between two electrodes is contained in every applicable theory, and these dwarf waves have been investigated experimentally at some length by Potapenko$^{(2)}$.

It should be noted that while the harmonic wave-lengths of a parallel-wire circuit having one end closed are proportional to 1, $\frac{1}{3}$, $\frac{1}{5}$, etc., the wave-lengths corresponding to harmonics of the electron-motion are proportional to 1, $\frac{1}{2}$, $\frac{1}{3}$, etc., so that there will be very few occasions where oscillations of harmonically related frequencies can be simultaneously maintained with a fixed external circuit. The fundamental wave-length corresponding to the electron motion in the seventh region of figure 5, for example, would be about 550 cm. This is not even approximately equal to any resonant wave-length of the connected circuit, and hence oscillations of this wave-length could not be maintained.

The position of the plate given in table 1 for each region is that for maximum plate-current. This may introduce a slight error because many experiments have shown that maximum plate current does not in general correspond to maximum oscillation-intensity. The effect could only be appreciable in the two widest regions, however, and an inspection of the possible correction shows that the analyses would not be affected.

Table 1. The grid voltage is constant at 144·2 V., and the cathode-grid distance at 5·5 mm.

Oscillation region	Optimum grid-plate distance (mm.)	Corresponding cathode-plate distance (mm.)	Wave-length λ (cm.)	$\frac{\lambda\sqrt{E_g}}{2000}$ (mm.)	Probable harmonic order	Space in which oscillations are probably located
1	7·0	12·5	120·5	7·2	Fundamental	Grid-plate
2	9·8	15·3	255·5	15·4	Fundamental	Cathode-plate
3	<9·9	<15·4	140·0	8·4	Fundamental	Grid-plate
4	14·8	20·3	182·5	10·6	1st Harmonic	Cathode-plate
5	18·7	24·2	416·0	24·9	Fundamental	Cathode-plate
6	21·8	27·3	252·0	15·1	1st Harmonic	Cathode-plate
7	24·1	29·6	182·5	10·5	2nd Harmonic	Cathode-plate

A large number of experimental curves similar to figure 5 have been obtained for different grid-potentials, circuit-lengths, grid-currents, and cathode-grid distances, and all submit to similar analysis leading to the same general conclusions. Figure 6 shows such a curve taken with a total circuit-length about one-tenth of that used for figure 5. The grid-potential and cathode-grid distance are approximately the same as before, but the grid current is about 10 per cent of the space-charge saturation value. The grid-plate capacity in this case forms a very appreciable part of the external circuit, and moving the plate through 20 mm. causes the fundamental wave-length of the net circuit to vary by nearly 30 cm. The slight rise in wave-length across the second and third regions will not be discussed here, but it is obvious that the mean wave-length of the regions decreases as the grid-plate capacity is decreased. The analysis of this curve is given in table 2. It is essentially the same as that for figure 5. The production of such a small number of regions is due to the fact that the short connected circuit has only one resonant wave-length at which oscillations might be expected to occur. The first harmonic wave-length of the circuit is 44 cm., and oscillations of this wave-length could be generated at the potential used only as dwarf waves of 4th or higher order.

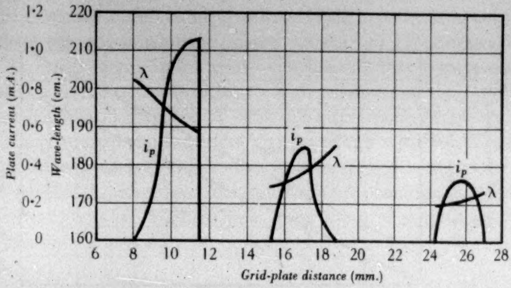


Figure 6. Oscillation regions obtained with the plane-electrode valve connected to a short external circuit, on varying the grid-plate distance. All other variables are held constant throughout. External circuit-length from terminating condenser to centre of cathode, 32·7 cm.; grid voltage, 146·5 V.; grid current, 6·0 mA.; cathode-grid distance, 4·8 mm.

Table 2. The grid potential is constant at 146·5 V. and the cathode-grid distance at 4·8 mm.

Oscillation region	Optimum grid-plate distance (mm.)	Corresponding cathode-plate distance (mm.)	Wave-length λ (cm.)	$\frac{\lambda\sqrt{E_g}}{2000}$ (mm.)	Probable harmonic order	Space in which oscillations are probably located
1	11·5	16·3	188·5	11·4	Fundamental	Grid-plate
2	17·0	21·8	178·5	10·8	1st Harmonic	Cathode-plate
3	25·7	30·5	171·0	10·3	2nd Harmonic	Cathode-plate

§ 6. RELATIONS BETWEEN WAVE-LENGTH AND THE BASIC PARAMETERS

The experiments of § 5 have shown that the wave-length of electron oscillations, where an external circuit is used, depends on three primary parameters. These are circuit-length, inter-electrode distance, and grid potential. It has been assumed that wave-length is related linearly to the first two of these and is inversely proportional to the square root of the grid potential. This assumption has been shown to be justified at a number of isolated values of the parameters.

Very simple experiments serve to show that these relations hold accurately over wide continuous ranges of all three parameters. The curve of figure 7 is obtained by measuring wave-length at several values of circuit-length, with plate-position adjusted for maximum plate current at each circuit-length. The circuit is connected to the grid and plate of the valve. Grid-potential, grid-current, and cathode-grid distance are constant throughout. The curve proves the linearity of the relations between wave-length and circuit-length, and between wave-length and inter-electrode distance, as exactly as the accuracy of experimental measurement will permit.

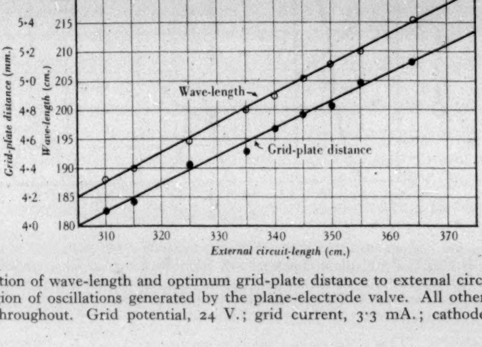


Figure 7. Relation of wave-length and optimum grid-plate distance to external circuit-length for a typical region of oscillations generated by the plane-electrode valve. All other variables held constant throughout. Grid potential, 24 V.; grid current, 3·3 mA.; cathode-grid distance, 5·5 mm.

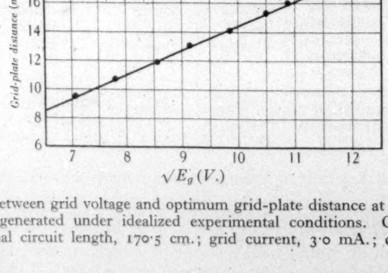


Figure 8. Relation between grid voltage and optimum grid-plate distance at constant wave-length, for oscillations generated under idealized experimental conditions. Constant wave-length, 235 cm.; external circuit length, 170·5 cm.; grid current, 3·0 mA.; cathode-grid distance, 8·4 mm.

By keeping the circuit-length and hence the wave-length constant, an inter-relation between grid potential and electrode spacing for optimum conditions can be obtained, as in figure 8. The cathode-temperature is low enough to prevent the grid current from becoming space-charge-saturated at any of the grid potentials used. The plate-position is adjusted for maximum oscillation-intensity, measured on a crystal circuit, at each point. The assumed relation between the parameters is again proved within experimental accuracy.

§7. OSCILLATION MECHANISM AND THEORY

It seems necessary to regard the two types of oscillation described in §§4, 5 as produced by distinctly different mechanisms. This hypothesis is not really a new one, but there has not previously existed satisfactory experimental evidence to justify it. The experiments in §4 prove conclusively that a space-charge-limited current flow in a triode whose grid is at a high positive potential, and whose plate is at cathode potential, is capable of radiating high-frequency energy without the assistance of any resonant inductance-capacity circuit whatever. They also prove that the oscillation period of this radiation corresponds to some electron transit time in the valve, and hence that the oscillations are characteristic of the electron-motion. The radiation is obtained only when the grid current is space-charge-limited. These oscillations may be designated by the name *space-charge electron oscillations*.

The oscillations at very low grid current dealt with in §5 are produced only when a resonant circuit is connected to the electrodes, and the oscillation wave-length is fixed at some harmonic wave-length of this circuit. The grid current need only be sufficient to supply a certain minimum amount of energy to the circuit, and may be a fraction of 1 per cent of the space-charge-limited value. The function of the electron-motion can only be the production of a negative resistance between the electrodes of the valve, to neutralize the positive resistance of the resonant circuit. These oscillations may be designated by the name *negative-resistance electron oscillations*.

It is not the intention here to develop any mathematical theory for the oscillations. There are so many variables involved that this must inevitably be a very difficult problem. As far as the space-charge oscillations are concerned, some theory analogous to those of Llewellyn and Benham may eventually be developed, but to explain the present experimental results it is necessary to consider two parameters which have not previously been included in space-charge theories. These are the grid-transparency and the distribution in phase of the electrons returning from the plate to the cathode-grid space. Only variation of the latter can account for the change in intensity of space-charge oscillations in the grid-cathode space observed on varying the plate-position.

With regard to the negative-resistance oscillations, Gill and Morrell(16) showed many years ago by very simple mathematical treatment that negative resistance may exist between the grid and plate of a positive-grid triode at frequencies approximately resonant fundamentally or harmonically with the electron transit time between these two electrodes. Alfven(17) has shown by much more thorough analysis involving all possible parameters that negative resistance may also exist between cathode and plate at frequencies similarly related to this longer transit time.

It may therefore be considered that there does exist some theoretical explanation of the oscillation regions of figure 5 and figure 6. The numerical evaluation of the results of Alfven's theory is unfortunately quite impracticable, so that direct comparison of theory with experiment is not possible. No other theory is sufficiently complete to give such comparison any significance. Negative-resistance theories all suggest that maximum oscillation-intensity does not in general occur when the oscillation-period is exactly equal to some electron transit time, but rather that the optimum ratio of these two may vary on either side of unity by several per cent, as a function of a number of the involved variables. It is questionable, however, whether indeterminacies in the electron motion can ever be sufficiently eliminated to enable this implication of the theories to be tested experimentally.

§ 8. ANALYSIS OF MORE COMPLEX RESULTS

It is desired to draw only three general conclusions from the experimental results given in this paper. The first of these is the necessity of recognizing two distinct oscillation mechanisms, as outlined in the last section. The second is that the period of electron oscillations generated in a triode under idealized experimental conditions is always very nearly equal to the electron transit time between one pair of the three electrodes. The third is that it is possible to obtain oscillations of period corresponding to each of the three possible inter-electrode transit times.

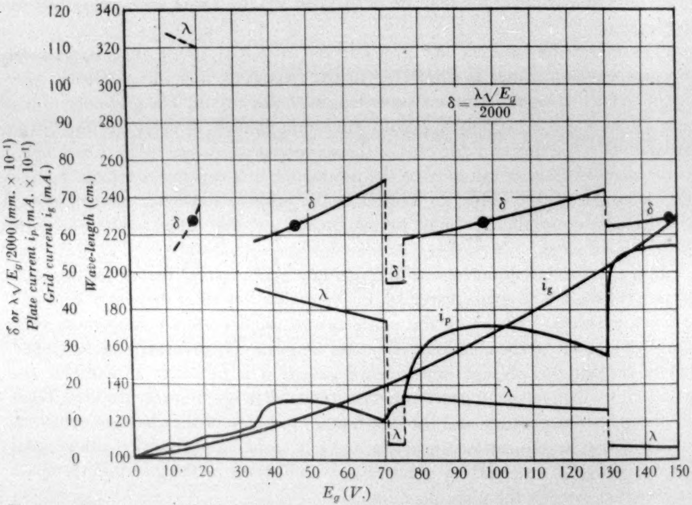


Figure 9. A more complex experimental curve that can be explained on the basis of the conclusions drawn from idealized experiments. All variables except grid potential are kept constant throughout. External circuit-length, 230 cm.; cathode-grid distance, 6·4 mm.; grid-plate distance, 4·9 mm. The black circles mark the value of δ, defined as $\lambda\sqrt{E_g}/2000$, at maximum plate current in each oscillation region. These are the points used for the analysis in table 3.

On the basis of these conclusions drawn from simplified experimental results it is possible to analyse much more complex results, from which the same conclusions could not have been initially deduced. Figure 9 is an example of these more complex results. The experimental conditions under which it has been obtained are a combination of those used in §§ 4, 5 above, in that an external circuit is connected to the grid and plate *and* the grid current is space-charge-limited. Electrode-spacings and cathode-temperature are kept constant throughout. The analysis of the five oscillation regions is given in table 3.

Table 3. The cathode-grid distance is constant at 6·4 mm., and the grid-plate distance at 4·9 mm., so that the cathode-plate distance is 11·3 mm.

Oscillation region	Optimum grid potential (V.)	Wave-length λ (cm.)	$\frac{\lambda\sqrt{E_g}}{2000}$ (mm.)	Probable harmonic order	Space in which oscillations are probably located
1	15	325	6·3	Fundamental	Cathode-grid
2	45	185	6·2	Fundamental	Cathode-grid
3	75	108	4·7	Fundamental	Grid-plate
4	97	131	6·4	Fundamental	Cathode-grid
5	142	106	6·3	Fundamental	Cathode-grid

Although the external circuit is connected to grid and plate, the analysis suggests that the oscillations in four of the regions are taking place between grid and cathode. A negative resistance explanation cannot be applied, and it must be concluded that the oscillations in these four regions have their source in the grid-cathode space charge, and are in effect space-charge oscillations, their wave-length being fairly narrowly controlled by the closely coupled low-resistance external circuit. The fifth region seems to be due to negative resistance in the grid-plate space. Since the grid-plate distance is smaller than the grid-cathode distance, and the grid-transparency is only 67 per cent, the total current in the grid-plate space can only be a small fraction of that required for space-charge limitation, possibly 10 to 20 per cent.

This experiment shows that if a circuit having a resonant wave-length λ is coupled in some way to a valve generating space-charge oscillations of the same wave-length λ, currents are induced in the circuit which react on the oscillator, and the oscillation-intensity is in general increased. Analogous effects are obtained with any oscillator and a closely coupled resonant circuit. Oscillations obtained under these conditions may be designated by the name *resonant space-charge oscillations.*

The blotting out of negative-resistance oscillations by the resonant space-charge oscillations in this experiment suggests that the latter are produced more efficiently than the former. This view is supported by the fact that it is found that resonant space-charge oscillations can be produced at much lower energy-inputs to the valve than negative-resistance oscillations can. In an extreme case these resonant space-charge oscillations have been observed at a grid potential of 4·2 V., with a grid current of 0·2 mA., making an input of 0·8 mW. It is important in relation to § 3 that this grid voltage is below the lowest ionization or resonance potential of any gas or vapour that might be considered present.

§ 9. CONCLUSIONS

In this paper the effect of continuous variation has been investigated only for the three primary parameters of grid potential, inter-electrode distance and length of external circuit. In addition to the primary parameters there are three minor variables which have a more or less critical effect on electron-oscillation mechanism. These are grid-transparency, plate potential, and space-charge density. In the present work the effects of the last two have been respectively eliminated by keeping the plate always at cathode potential and by using either maximum or negligible space-charge density. With regard to the negative-resistance oscillations, further experiment indicates that grid-transparency has not a very critical effect on either wave-length or oscillation-intensity. The results of figure 6 can be duplicated almost exactly with grids of transparency up to 90 per cent. Space-charge oscillations, however, are considerably affected by this variable. The results shown in figure 3 have been repeated for three different grids of approximately the same transparency, 67 per cent, but the oscillations produced with a grid-transparency of 80 per cent, for example, are no longer confined to the cathode-grid space. It is hoped in a future paper to deal in detail with this effect, and also with the results obtained by continuously varying each of the other minor parameters.

How far the results of the present work can be applied to experiments on cylindrical-electrode valves is not clear. There seems never to have been any indication that oscillations obtained from commercial valves can be restricted to the grid-filament space or to the grid-plate space as has been found here. The critical effect of heating-current on the space-charge oscillations generated by a commercial valve, figure 4, also suggests an inherent difference between the oscillation mechanisms for plane and cylindrical electrodes.

§ 10. ACKNOWLEDGMENTS

It is a pleasure to acknowledge my indebtedness to Mr J. A. Ratcliffe for many helpful suggestions, and to thank him for his continued interest in this work. I wish also to thank the McGill University chapter of the Delta Upsilon fraternity for a scholarship which enabled this work to be started.

REFERENCES

(1) Megaw, E. C. S. *J. Instn elect. Engrs*, **72**, 313 (1933).
(2) Potapenko, G. *Phys. Rev.* **39**, 625 (1932).
(3) Moore, W. H. *J. Franklin Inst.* **217**, 347 (1934).
(4) Chapman, F. W. *Wireless Engr*, **8**, 500 (1932).
(5) Nettleton, L. L. *Proc. nat. Acad. Sci.*, Wash., **8**, 353 (1932).
(6) Tonks, L. and Langmuir, I. *Phys. Rev.* **33**, 195 (1929).
(7) Grechowa, M. T. *Z. Phys.* **35**, 59 (1926).

(8) Rindfleisch, H. *Ann. Phys.*, Lpz., **14**, 273 (1932).
(9) Barkhausen, H. and Kurz, K. *Phys. Z.* **21**, 1 (1920).
(10) Megaw, E. C. S. *J. Instn elect. Engrs*, **72**, 326 (1933).
(11) McPetrie, J. S. *Phil. Mag.* **16**, 284 and 544 (1933).
(12) Kapzov, N. and Gwosdower, S. *Z. Phys.* **45**, 114 (1927).
(13) Moore, W. H. *Proc. Inst. Radio Engrs*, N.Y., **22**, 1021 (1934).
(14) Benham, W. E. *Phil. Mag.* **11**, 496 (1931).
(15) Llewellyn, F. B. *Proc. Inst. Radio Engrs*, **21**, 1532 (1933).
(16) Gill, E. W. B. and Morrell, J. H. *Phil. Mag.* **49**, 369 (1925).
(17) Alfven, H. Uppsala University Dissertation (1934).

546.11.02

THE EFFICIENCY OF SEPARATION OF HYDROGEN AND DEUTERIUM BY ELECTROLYSIS

By T. H. ODDIE, M.Sc., Natural Philosophy Laboratory,
University of Melbourne

Communicated by Prof. T. H. Laby, May 10, 1935

ABSTRACT. A method of determining the efficiency of concentration of D_2O by electrolysis of mixtures of H_2O and D_2O is described; it takes account of the effects of losses by evaporation and spraying. It is found that the hydrogen-ion concentration of the electrolyte and the nature of the electrodes do not affect the efficiency α, but α increases with increasing current-density. The coefficient α is found to be $4{\cdot}0 \pm 0{\cdot}2$ with current-density 0·6 A./cm.², and $4{\cdot}6 \pm 0{\cdot}1$ with current-density 2 A./cm.² These results are in agreement with values predicted theoretically by Urey and Teal.

§ 1. INTRODUCTION

Experimental work hitherto published shows a marked uncertainty not only of the value of the coefficient of separation of hydrogen and deuterium by the electrolysis of water under given conditions, but also of the effect on the coefficient of changes in these conditions; thus, published values vary between 2·6 and 7·3. Bell and Wolfenden⁽¹⁾ found that the separation coefficient α depends solely on current-density, being greatest for high current-densities, while Collie⁽²⁾ maintains that for efficiency a current-density less than 0·1 A./cm.² is necessary. On the other hand, Topley and Eyring⁽³⁾ found a dependence of α upon the nature of the electrodes and upon the hydrogen-ion concentration of the electrolyte. Collie⁽²⁾ also obtained different values of α with different experimental conditions.

Values of α obtained by other workers are given in numerical order in table 1, together with the nature of the electrodes and electrolyte, the current-density and the mean D_2O concentration. A study of this table fails to indicate any certain dependence of α on any of the above factors, but it should be borne in mind that in some cases experimental errors are admitted to be large, and apparently no correction has been made for the inevitable loss of water by evaporation and spraying during electrolysis.

In this paper is described an attempt to measure α with greater certainty and to study the effects on α of four factors, the current-density, the nature of the electrodes, the hydrogen-ion concentration (pH) of the electrolyte and D_2O concentration.

§ 2. THEORY OF THE EXPERIMENT

When a mixture of D_2O and H_2O is electrolysed, the proportion of D_2 molecules to H_2 in the evolved hydrogen gas is less than the ratio of D_2O to H_2O in the electro-

Table 1

Reference number	Electrolyte	Electrodes	Current-density A./cm.²	Concentration of D_2O	α
2	H_2SO_4	Pb	0·1	0·002	2·6
1	NaOH	Ni	0·07	0·001	3·7
1	NaOH	Ni	—	0·004	4·0
1	NaOH	Ni	—	0·002	4·8
13	NaOH	Ni	—	—	5·0
3	KOH	Ag	1	0·07	5·2
1	NaOH	Pt	—	0·002	5·3
1	NaOH	Cu	—	0·002	5·3
3	KOH	Ni	1	0·07	5·5
3	H_2SO_4	Cu	1	0·07	5·6
1	NaOH	Ni	10	0·001	5·6
3	H_2SO_4	Pt	1	0·07	5·7
4	NaOH	Ni	0·7	—	5·8
3	KOH	Cu	1	0·07	6·8
3	KOH	Pt	1	0·07	7·0
3	KOH	Fe	1	0·07	7·2
3	KOH	Pb	1	0·07	7·3

lyte. A separation coefficient α may be defined as follows. Let a small fraction of the water be decomposed electrolytically; then the fraction of H_2O molecules decomposed will be α times that of D_2O molecules.

α

The relation between α and the initial and final volumes and D_2O-concentrations of the water has been given by Harteck⁽⁴⁾, but losses by spraying and evaporation of the solution during electrolysis modify the theoretical equation which he gives as is shown below.

Let N_0 be the number of molecules of H_2O in water initially; N the number of molecules of H_2O in water at time t; n_0 the number of molecules of D_2O in water initially; n the number of molecules of D_2O in water at time t; K the total number of molecules electrolysed per second; and k the total number of molecules evaporated per second.

N_0, N
n_0
n, t, K
k

Now the vapour pressure of D_2O is about 0·9 times that of H_2O⁽⁵⁾, but for a correction for loss by evaporation the assumption that the vapour pressures are equal gives a sufficiently close approximation. Now for a small loss by electrolysis,

$$\frac{(dN)_1}{(dn)_1}=\frac{N}{n}\alpha \qquad \ldots\ldots(1\cdot11),$$

and for a small loss by evaporation,

$$\frac{(dN)_2}{(dn)_2}=\frac{N}{n} \qquad \ldots\ldots(1\cdot12),$$

where the suffixes 1 and 2 imply electrolysis and evaporation respectively. Further, the number of molecules lost in time dt is given by

1, 2

$$(dN)_1+(dn)_1=-K\,dt \qquad \ldots\ldots(1\cdot21),$$

and

$$(dN)_2+(dn)_2=-k\,dt \qquad \ldots\ldots(1\cdot22).$$

From equations (1·11) and (1·21),

$$(dN)_1 = -\frac{KN\alpha}{N\alpha+n}\,dt \qquad \text{......(1·31)},$$

$$(dn)_1 = -\frac{Kn}{N\alpha+n}\,dt \qquad \text{......(1·32)},$$

and from equations (1·12) and (1·22),

$$(dN)_2 = -\frac{kN}{N+n}\,dt \qquad \text{......(1·41)},$$

$$(dn)_2 = -\frac{kn}{N+n}\,dt \qquad \text{......(1·42)}.$$

Thus the total number of molecules lost in time dt is

for H_2O, $$dN = (dN)_1 + (dN)_2 = -\frac{KN\alpha}{N\alpha+n}\,dt - \frac{kN}{N+n}\,dt \qquad \text{......(1·51)},$$

and for D_2O, $$dn = (dn)_1 + (dn)_2 = -\frac{Kn}{N\alpha+n}\,dt - \frac{kn}{N+n}\,dt \qquad \text{......(1·52)}.$$

These equations (1·51) and (1·52) may be combined and integrated to give a relation

$$\frac{\alpha-1}{1+r}\log\frac{N+n}{N_0+n_0} = \log\frac{1-C}{1-C_0} - \alpha\log\frac{C}{C_0} \qquad \text{......(1·6)},$$

r, C where $r = k/K$ and $C = n/(N+n)$ = the concentration of D_2O.

This relation is similar to that of Harteck but has the added factor $1/(1+r)$ and the ratio of the initial and final numbers of molecules $(N+n)/(N_0+n_0)$ is substituted for the ratio of the initial and final volumes.

W_0, W The ratio $(N+n)/(N_0+n_0)$ is connected with the ratio of initial and final masses W_0, W of water by the relation

$$\frac{W}{W_0} \doteq \frac{N+n}{N_0+n_0}\left\{1+\frac{C-C_0}{9}\right\}$$

C_0 very nearly when n/N_0 is small, where C_0 is the initial concentration of D_2O.

§ 3. EXPERIMENTAL PROCEDURE

Conditions for accuracy of α. By differentiating equation (1·6) a relation between the errors in W/W_0 and C/C_0 and the resulting error in α may be found. The error in r may be neglected, and also the first term on the right-hand side of the equation. We obtain

$$\frac{d\alpha}{\alpha} = \left[\frac{1}{\log W/W_0} - \frac{1}{\log W/W_0 + (1+r)\log C/C_0}\right]\frac{d\,(W/W_0)}{W/W_0} - \left[\frac{1}{\dfrac{\log W/W_0}{1+r} + \log C/C_0}\right]\frac{d\,(C/C_0)}{C/C_0} \qquad \text{......(2·1)},$$

$$= 1\cdot6\,\frac{d\,(W/W_0)}{W/W_0} - 2\cdot2\,\frac{d\,(C/C_0)}{C/C_0} \qquad \text{......(2·2)},$$

after substitution of the experimental values given below.

Thus 1-per-cent errors in W/W_0 or in C/C_0 give considerably larger errors in α.

Conditions of experiment. The influence of four factors, (i) current-density, (ii) the hydrogen-ion concentration of the electrolyte, (iii) the nature of electrodes, and (iv) the concentration of D_2O, has been studied, one of these factors being varied at a time.

The experiments were carried out in water-cooled test tubes with either wire or metal-strip electrodes. In some the evaporation losses were found by calculation from the initial and final volumes of water and the input of electrical energy. In the more accurate experiments, however, the latter quantity was not determined, but losses were measured directly by passing the electrolytic gases through a glass vessel cooled with solid carbon dioxide, and weighing the condensed liquid. The concentration of D_2O was measured by a float method described in a later section.

In a typical experiment the initial concentration of the electrolyte, sodium hydroxide, was 1·4 per cent. The concentration of D_2O was from 2·5 to 10 per cent. The current-density was high—about 2 A./cm.^{2} The electrodes were of nickel wire. The mass of electrolyte was 1·895 g. The masses W_0, W were 133·695 g. and 19·769 g. respectively. Hence it was calculated that the mass of water evaporated was 7·407 g. and the mass of water electrolysed was 106·519 g. The concentration of D_2O in the tap water was 0·0001 per cent; the initial concentration C_0 was 0·0246 ± 0·00002, and the final concentration C was 0·0970 ± 0·0003. The corresponding equilibrium temperatures of the float were 17·21 ± 0·005, 26·635 ± 0·005, and 49·5 ± 0·1° C. Hence from equation (6) α was found to be 4·24 ± 0·05.

§ 4. EXPERIMENTAL RESULTS

Hydrogen-ion concentration of electrolyte. The current-density was about 0·2 A./cm.^{2} Electrodes of lead were used with acid solutions, and of stainless steel with alkaline solutions. The mean concentration of D_2O was 0·002.

Table 2

Number of experiments	Electrolyte	α
3	H_2SO_4	2·9 ± 0·3
13	NaOH	3·1 ± 0·3

Thus the hydrogen-ion concentration of the electrolyte does not appear to affect α.

Nature of electrodes. The current-density was about 0·6 A./cm.^{2} The electrolyte was sodium hydroxide, from 1 to 6 per cent. The mean concentration of D_2O was 0·002.

Table 3

Number of experiments	Electrodes	α
1	Monel	3·98 ± 0·2
3	Nickel	3·93 ± 0·2

α is independent of the electrodes, as is also shown by the experiment at a lower current-density with lead and stainless-steel electrodes given in table 2.

Concentration of D_2O. The current-density was about 0·2 A./cm.² The electrodes were stainless steel or iron. The electrolyte was sodium hydroxide.

Table 4

Number of experiments	Concentration of D_2O	α
17	0·0015	3·1 ± 0·3
11	0·0045	2·9 ± 0·3

α is a constant for different stages in the process of concentration.

Current-density. The electrodes were of nickel, iron, stainless steel and nickel-chromium, α being independent of these. The electrolyte was sodium hydroxide. The mean concentration of D_2O was 0·002 to 0·06.

Table 5

Number of experiments	Current-density (A./cm.²)	α
22	Low, about 0·2	3·0 ± 0·3
3	Medium, about 0·6	4·0 ± 0·2
3	High, about 2	4·6 ± 0·05

It is evident that the current-density has a marked influence on α, high current-densities being the most efficient. The determination of the actual value of the current-density was rendered very uncertain by such factors as varying depth of the electrolyte, current-variations, the shape of the coiled-wire electrodes, and the uncertain conducting surface. The electrodes employed were as follows. For low current-densities, metal strips 2 cm. wide, in 1-in. test tubes; for medium current-densities, metal strips 1 cm. wide, in ½-in. test tubes; and for high current-densities, metal wires in concentric coils.

Value of α. The eight most accurate results for α are shown in table 6.

Table 6

Current-density (A./cm.²)	0·6	2
α	3·98 ± 0·2 4·30 ± 0·2 3·65 ± 0·2 3·85 ± 0·2 Mean 4·0 ± 0·2	4·84 ± 0·2 5·12 ± 0·08 4·24 ± 0·05 4·48 ± 0·08 Mean 4·6 ± 0·1

The differences between these readings exceed the experimental errors and are due to the variations of current-density, experiments having been performed in different test-tubes with different coiled-wire electrodes.

§ 5. MEASUREMENT OF CONCENTRATION OF D_2O

The molecular concentration of D_2O in the D_2O-H_2O mixtures of the above experiments was determined with a cylindrical pyrex float 1·5 cm. long and 0·5 cm. in diameter. This was so constructed as to float in ordinary water at 17·21° C. The difference between the temperature at which the float neither rose nor sank when completely immersed in ordinary water, and a similar temperature for the mixture under observation, gave a measure of the difference of D_2O concentrations.

The concentration of D_2O in tap water was taken as 1 in 10,000(6), which figure agrees with a determination of 1 in 8500 made in this laboratory by Christiansen, Crabtree and Laby, but not with earlier estimates(7).

Water samples were distilled five times, a tin condenser being used for all but the first distillation. A small correction was necessary for the thermal expansion of the float, but calculation showed that the effect of changes in the volume of the float with variations of barometric pressure was negligible.

The water sample was placed in a narrow test tube containing the float, and the whole was immersed with a Beckmann thermometer in a large well-stirred water bath, the temperature of which was varied slowly. Small movements of the float were easily followed by projecting a magnified image of the apparatus on to a millimetre scale at a distance of about 6 ft.

The calculation of C was carried out by means of a formula derived as follows. Taking for the present ordinary distilled tap water as pure H_2O, let C' be the excess molecular concentration of D_2O in the mixture over that in the tap water, K_1 a constant, d_D the density of pure D_2O, and d_H that of H_2O. C' K_1, d_D, d_H

Since the molecular weight of $D_2O=20$ and of $H_2O=18$, the total mass of water is

$$K_1\,(20n+18N) \qquad \ldots\ldots(3{\cdot}11),$$

and the total volume of water is

$$K_1\left(\frac{20n}{d_D}+\frac{18N}{d_H}\right) \qquad \ldots\ldots(3{\cdot}12),$$

assuming that when H_2O and D_2O in the liquid state are mixed the volume of the mixture is the sum of the volumes of the constituents.

Hence the mean density

$$\rho=\left(\frac{20n+18N}{20nd_H+18Nd_D}\right)d_H\,d_D \qquad \ldots\ldots(3{\cdot}21), \qquad \rho$$

$$=\left(\frac{n+0{\cdot}9N}{nd_H+0{\cdot}9Nd_D}\right)d_H\,d_D \qquad \ldots\ldots(3{\cdot}22).$$

But $$C'=n/(N+n) \qquad \ldots\ldots(3{\cdot}31).$$

Therefore $$n=C'N/(1-C') \qquad \ldots\ldots(3{\cdot}32).$$

Thus from equations (3·22) and (3·32)

$$\rho=\frac{\left(\dfrac{C'}{1-C'}+0{\cdot}9\right)d_H\,d_D}{\left(\dfrac{C'}{1-C'}\right)d_H+0{\cdot}9\,d_D} \qquad \ldots\ldots(3{\cdot}41).$$

Whence
$$C'\left[10\left\{\frac{1}{d_H}-\frac{1}{d_D}\right\}-\left\{\frac{1}{d_H}-\frac{1}{\rho}\right\}\right]=9\left\{\frac{1}{d_H}-\frac{1}{\rho}\right\} \qquad \ldots\ldots(3{\cdot}42).$$

But ρ is the density of the float at equilibrium temperature t in the mixture. Thus $\rho=\rho'\{1-\beta(t-t')\}$, where ρ' is the density of the tap water at equilibrium temperature and t' that of the float in tap water, and β is the volume coefficient of thermal expansion of the float.

ρ'
t', β

Finally C, the true concentration of $D_2O = C' + 0{\cdot}0001$, since tap water has already one part in 10,000 of D_2O.

C could thus be calculated since ρ, d_H, d_D were known and β ($3{\cdot}1\times10^{-6}$) was measured by determining the resting-point of the float in a solution of sodium chloride of known density.

§ 6. CONCLUSIONS

The experimental results given indicate that α, the efficiency of concentration, depends solely on the current-density, a conclusion which is in agreement with those of Bell and Wolfenden(1), but not with those of Collie(2) or Topley and Eyring(3).

The actual value, 4·6, of α obtained with high current-density is lower than the mean, 5·4, of values given in table 1, but is in better agreement with the theoretically predicted values.

The coefficient α varies with temperature(8), but results given in this paper and in table 1 all relate to room-temperature.

The theoretical deduction of the value of α is complicated by the fact that equilibrium conditions do not hold in practice during electrolysis, and by the necessity of making various simplifying assumptions. The equilibrium constant of the reaction

$$H_2 + HDO = HD + H_2O$$

leads(9) to a value of α equal to 3 and equilibrium conditions have been obtained approximately by Washburn, Smith and Frandsen(10), who found values of α of 1·49, 2·71, 3·14, 2·81 and 2·79.

α has been derived theoretically by Bell(8) and by Topley and Eyring(11), who gave a value of about 20; their theory has been criticized by Urey and Teal(12), who calculate values 4·0, 4·3 and 4·55 at 0, 50 and 100 moles per cent of D_2O respectively. These are in better accord with experiment.

§ 7. ACKNOWLEDGMENTS

The writer has been in receipt of a research grant from the University of Melbourne while he was engaged on this investigation, the expenses of which were provided by a gift to this University from the Associated Smelters Company of Australia. Mr Germann, of the Melbourne and Metropolitan Tramways Board, very generously made available the electrolyte contained in alkaline batteries which had become enriched in deuterium. The writer wishes to thank Prof. T. H. Laby, F.R.S., for suggesting the investigation and for his valuable help throughout.

REFERENCES

(1) Bell and Wolfenden. *Nature*, Lond., **133**, 25 (1934).

(2) Collie. *Nature*, Lond., **132**, 568 (1933).

(3) Topley and Eyring. *J. Amer. chem. Soc.* **55**, 5058 (1933).

(4) Harteck. *Proc. phys. Soc.* **46**, 277 (1934).

(5) Lewis and McDonald. *J. Amer. chem. Soc.* **55**, 3057 (1933).

(6) Ingold, Ingold, Whitaker and Whytlaw-Gray. *Nature*, Lond., **134**, 661 (1934).

(7) Birge and Menzel. *Phys. Rev.* **37**, 1669 (1931).
Hardy, Barker and Dennison. *Phys. Rev.* **42**, 279 (1932).
Lewis and McDonald. *J. chem. Phys.* **1**, 341 (1933).
Tate and Smith. *Phys. Rev.* **43**, 672 (1933).
Bleakney and Gould. *Phys. Rev.* **44**, 265 (1933).

(8) Urey. *Science*, **78**, 566 (1933).
Bell. *J. chem. Phys.* **2**, 164 (1934).
Guntherschultze and Keller. *Z. Elektrochem.* **40**, 182 (1934).
Schwartz, Küchler and Steiner. *Z. Elektrochem.* **40**, 298 (1934).

(9) Crist and Dalin. *J. chem. Phys.* **2**, 442, 548 (1934).
Farkas and Farkas. *J. chem. Phys.* **2**, 468 (1934); *Proc. roy. Soc.* A, **146**, 623 (1934).

(10) Washburn, Smith and Frandsen. *Bur. Stand. J. Res.* **11**, 453 (1933).

(11) Topley and Eyring. *J. chem. Phys.* **2**, 217 (1934); *J. Amer. chem. Soc.* **55**, 5058 (1933).

(12) Urey and Teal. *Rev. mod. Phys.* **7**, 34 (1935).

(13) Lewis and McDonald. *J. chem. Phys.* **1**, 341 (1933).

535.62

A TRICHROMATIC COLORIMETER

By R. DONALDSON, M.A., Optics Division,
National Physical Laboratory

Received March 15, 1935

ABSTRACT. A trichromatic colorimeter is described in which the integrating properties of a diffusing sphere are employed to effect the mixture of the three instrumental stimuli. Means are described for obtaining linearity of the scales. The general construction has been planned with a view to keeping the design as simple as is consistent with efficiency.

---

§ 1. INTRODUCTION

Routine practice of colour-measurement for scientific and industrial purposes makes considerable demands on the measuring-instrument in the way of such requirements as accuracy, simplicity of construction, ease of maintenance. At present these demands seem to be most adequately met by the Guild trichromatic colorimeter(1) manufactured by Messrs Adam Hilger Ltd. For various reasons it is thought the production of another colorimeter of equal accuracy and efficiency would be of value. The Guild colorimeter is an expensive instrument. The original model was designed and made for research work at the National Physical Laboratory, where special supplies of direct current can be obtained as required. There is some difficulty, however, in adapting it for use with alternating current, on account of the method of colour-mixing employed. In view of the progressive change from d.c. to a.c. in the public supply throughout the country, this disadvantage is becoming of increasing importance. The present instrument can be used equally well with d.-c. or a.-c. supply. The cost of construction should be considerably less than that of the Guild instrument, and it should therefore find a field of utility wider than that which was available when the Guild colorimeter alone could be considered an instrument of high accuracy.

The general principle underlying the design of visual trichromatic colorimeters may be briefly summarized as follows. Provision is made for the coloured light from the specimen under examination to fill one half of a bipartite photometric prism. The other half is filled by a mixture of the three matching-stimuli. These matching-stimuli or working primaries are colours of high saturation, usually a red, a green and a blue. Means are also provided to vary independently and in a known manner the intensities of the matching-stimuli. The basis of operation consists in finding the intensities of these stimuli which give a colour-match with the specimen.

As attention here is confined to the construction of the instrument only, for particulars regarding method of use, transformation of results to standard form, etc. reference should be made to the paper quoted above(1) and also to the paper on the C.I.E. colorimetric system by Smith and Guild(2). The trichromatic colorimeters which have been described from time to time differ only in the methods by which the matching-stimuli are produced, mixed, and varied in intensity. In the Guild instrument coloured gelatine filters illuminated by a gas-filled lamp constitute the matching-stimuli. The mixing of these is effected by a rotating prism which transmits each stimulus successively to the field of view with sufficient rapidity for no flicker, either of colour or of brightness, to be observed. The method of variation of intensities can be regarded as an application of the principle of the variable rotating sector, but instead of the customary combination of a rotating sector and a stationary beam of light, a prism which rotates the beam of light and a stationary sector are used. In the instrument now to be described the chief feature of the design is the replacement of the rotating prism by a diffusing integrating sphere.

§ 2. CONSTRUCTION OF THE INSTRUMENT

Figure 1 shows the plan of the instrument and figure 1*a* an end elevation showing the plate *A* which faces the light-source.

The light-source *L* is a 250-W. lamp of the projector type run at 90 per cent of its rated volts. This reduction of voltage ensures longer life and greater uniformity in emission of radiation. A source of this rating at least is needed if a sufficiently high field-illumination is to be obtained.

The three rectangular apertures *R*, *G*, *B* in the plate *A* can be uncovered to varying extents by shutters sliding in V-shaped grooves. The dimensions of each aperture are 8·5 cm. × 2·0 cm. Each shutter is operated by a separate rack and pinion. A scale mounted on the shutter allows the extent of the opening to be read. The edges of the shutter and aperture are bevelled, and special attention has been paid to the fit when the shutter is on the point of closing.

Behind the apertures are mounted the colour filters, the red being placed in aperture *R* and so on. For reasons of permanence the red and blue filters are of glass. The red is a selenium red and the blue is Chance's No. 7 Contrast filter. As no suitable green glass seems to be available at present, the green Wratten filter as used in the Guild instrument has been retained.

The condenser lens *C* which follows is of the usual type. It was manufactured from plate glass of good quality; its diameter was 6 in. and the radius of curvature of the curved surfaces was also 6 in. The lens collects the light transmitted by the three apertures and brings it to a focus at the opening *D* in the sphere *S*.

The sphere is 6 in. in diameter, is made of copper, and is silver-plated inside. A layer of magnesium oxide is smoked on to the silver in the usual way. The apertures *D* and *F* are 1·3 in. and 0·25 in. respectively in diameter. Glass windows covering these apertures prevent the magnesium oxide from being contaminated with dust. The appropriate half of the photometric prism *P* is illuminated by light from the sphere-wall at *E* via opening *F* and prism *H*. A silvered mirror of thin glass can be substituted for the prism *H*. The function of lens *O* is to produce an

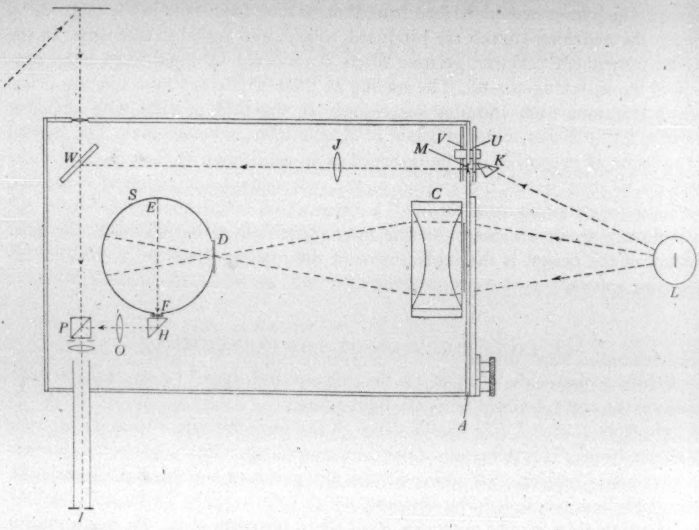


Figure 1.

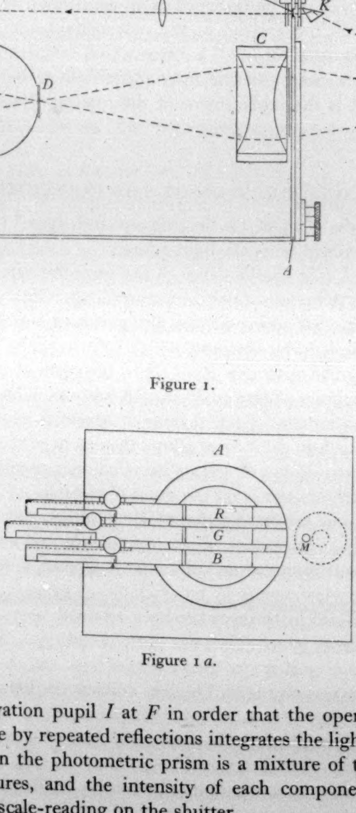


Figure 1 *a*.

image of the observation pupil *I* at *F* in order that the opening *F* may be kept small. As the sphere by repeated reflections integrates the light which enters at *D*, the light observed in the photometric prism is a mixture of the light transmitted by the three apertures, and the intensity of each component is approximately proportional to the scale-reading on the shutter.

In other respects, such as the method of obtaining the standard size of field and the optical system for viewing the specimen, the construction follows that employed in the Guild instrument. There is a slight variation in the arrangement for adding desaturating colour to the colour under test. The prism K acts as a mirror and deflects the light from L through M. In the small aperture there is mounted a piece of glass ground on both sides to act as a diffuser. The remainder of the system, comprising the plate U to hold the colour filters, the circular photometer wedge V, the lens J, and the glass plate W is similar to the corresponding part of the Guild instrument.

§ 3. LINEARITY OF THE SCALES

For a colorimeter to give a satisfactory performance the standard of accuracy required in the scales of the matching stimuli is such that, when calibrated against sectors by means of brightness-matches in the small photometric field, no measurable error should be found. A more sensitive test can be made by comparing the effectiveness of different small parts of the aperture with each other. Experience shows that variations up to 2 per cent from the mean value can be tolerated in this test. In the present design the factors which tend to destroy the proportionality between scale-reading and intensity of stimulus are as follows. (*a*) Uneven horizontal distribution of luminous flux within the cone subtended by the aperture at the source. (*b*) Uneven distribution of light over each aperture due to (i) varying distance from the lamp, (ii) varying obliquity. (*c*) Non-uniform transmission through the coloured filters. (*d*) Varying reflection and absorption losses in the condenser lens. (*e*) Varying effectiveness of the sphere as an integrator for pencils of light entering at different angles.

It should be noted that, on account of the fact that the shutters slide horizontally, vertical variations in the above quantities are of no consequence and only horizontal variations need be taken into account.

The magnitude of (*a*) was measured in the following manner. The projector lamp was mounted on a photometric bench and rotated through the required angle. A photometer head some distance away served as an indicator for variation in brightness. These are the conditions appropriate to the middle aperture. The upper and lower apertures require the projector lamp to be tilted during the measurements. In all cases no measurable variation could be detected. It is important that no obscuring of one filament by another should take place in the cone defined by the aperture. For this reason it is desirable to employ the type of lamp in which the filaments are all in one plane. The effect of (*b*) can be calculated from the inverse-square law together with the cosine law. For the middle aperture the variation amounts to 6 per cent.

To insist on absolute uniformity in (*c*) would involve the rejection of an excessive number of filters. Moreover, to draw up a calibration curve of the variation in transmission which is applicable when the filter is mounted in the instrument is rather difficult. The most practicable course is therefore to reserve the available

2-per-cent variation for this item alone, that is, to permit variations not exceeding 2 per cent in the local transmission through filters. Even with this latitude some difficulty has been experienced in finding a suitable piece of selenium glass for the red filter. To calculate (*d*), rays corresponding to certain points in the apertures were traced through the condenser lens, and the reflection losses at the surfaces calculated from Fresnel's formula and the thicknesses of glass traversed measured. To estimate the absorption losses the absorption per centimetre of the glass for light of the composition transmitted through the apertures was required. This was supplied by separate measurements on a plane slab of the same material as that from which the lens was manufactured. The variation due to these two sources combined was a maximum for the red aperture where it attained a magnitude of 7 per cent.

The following simple test gave a measure of the variation due to (*e*). A beam of light was allowed to enter the sphere and the brightness of a fixed point of the interior wall was observed when the sphere was rotated through an angle equal to the angle subtended at the opening of the sphere by an aperture containing a colour filter. No measurable alteration in brightness could be detected. Deterioration of the magnesium-oxide coating can only affect the performance of the instrument by destroying the integrating power of the sphere to such an extent that the above test would produce a positive result. As the interior is completely protected, experience with similar surfaces shows that no significant change is likely to take place in a shorter period than three months and the coating therefore need only be renewed at that interval.

By combining (*b*) and (*d*) a calibration curve was drawn for each aperture. It was found that the curve which had been calculated to refer to a line along the middle of the aperture was sensibly identical with those referring to the top and bottom portions and so could be applied to the whole aperture. As effects (*b*) and (*d*) work in opposite directions the final correction was much less than had been anticipated. In the case of the blue there was practically complete compensation; with the green the centre of the aperture was 2½ per cent more effective than the ends; with the red the ends were more effective than the centre by 1½ per cent. As the application of calibration corrections is undesirable when it can be avoided, the need for corrections has been eliminated by modifying the widths of the apertures by the insertion of templates, which compensate for these variations and so render the scale strictly linear. Figure 2 shows these templates exaggerated in size. It is advisable to fit them at both top and bottom so that each shares half of the correction because in this way any errors due to an irregular vertical distribution of light from the lamp are averaged out.

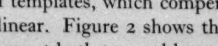


Figure 2. Rectangular aperture modified by templates.

It should be noted that on account of the arrangement of the spiral filaments in a projector lamp the vertical variations in brightness are considerably larger than the horizontal. This fact caused the abandonment of a design in which it was thought the need for calibration corrections would be eliminated by giving the apertures the form of sectorial openings.

A final test of the scales in the completed instrument, made by means of sectors, showed that they were sufficiently accurate for colorimetric measurements. Generally the departures from linearity were less than 1 per cent. In the case of very small openings in which readings cannot be made to this degree of accuracy, the scales were found to be true to the limit to which they could be read.

§ 4. PERFORMANCE

Throughout the design the aim has been to keep the construction simple so far as this could be done without loss of efficiency and, as far as possible, to employ only components which can be easily and cheaply obtained. The only high-grade optical part is the photometric prism. As no correction for chromatic or other aberrations is necessary, the small lenses are ordinary spectacle glasses. Of the metal parts only the apertures and shutters require high-grade workmanship.

The standard of performance of this colorimeter is equal to that of the Guild instrument.

§ 5. ACKNOWLEDGMENTS

In conclusion I should like to thank Mr J. Guild for his interest in the instrument and for the helpful suggestions he has made, and Mr Buxton for his advice on points in the mechanical design. The instrument was constructed in the Optics Workshop, National Physical Laboratory.

REFERENCES

(1) Guild, J. *Trans. opt. Soc.* **27**, 106 (1925–6).
(2) Smith, T. and Guild, J. *Trans. opt. Soc.* **33**, 73 (1931–2).

538.56:537.56 551.51.053.5

THE INFLUENCE OF A MAGNETIC FIELD ON THE HIGH-FREQUENCY CONDUCTIVITY OF AN IONIZED MEDIUM

By E. V. APPLETON, F.R.S.

AND

D. B. BOOHARIWALLA, M.Sc., Ph.D., A.Inst.P.,
Wheatstone Laboratory, King's College, London

Received May 21, 1935

ABSTRACT. The relation between the transverse high-frequency conductivity and the pressure of ionized air under the influence of an imposed magnetic field has been studied experimentally. The pressure at which such conductivity is a maximum is found to vary with the intensity of the magnetic field as is to be expected on theoretical grounds if the high-frequency conductivity is due solely to electrons. The significance of the results in connexion with ionospheric conductivity is discussed.

---

§ 1. INTRODUCTORY

In a previous communication⁽¹⁾ from this laboratory an account was given of an experimental investigation of the variation of the radio-frequency conductivity of ionized air with pressure. It was there shown that, from the value of the critical pressure at which such conductivity for a given ionization-content was found to be a maximum, the magnitude of the collisional frictional forces experienced by vibrating electrons could be estimated.

The present communication deals with an extension of such experiments to cases in which a magnetic field is imposed on the ionized medium. This extension was prompted by the need for checking certain deductions from the magneto-ionic theory⁽²⁾ of wireless transmission concerning the absorptive effect of upper-atmospheric conductivity on waves travelling in the ionosphere.

§ 2. THE RADIO-FREQUENCY CONDUCTIVITY OF AN IONIZED MEDIUM IN A MAGNETIC FIELD

The theory of the propagation of electromagnetic waves through an ionized medium under the influence of an imposed magnetic field was given by H. A. Lorentz as early as 1909⁽³⁾. From the results obtained by Lorentz we can deduce, with but slight modifications, the expression which we use here for the radio-frequency conductivity of the medium.

We consider the medium as consisting of N electrons per cm^3, each of charge e and mass m, the number of electrons being small compared with the number of uncharged molecules in a given volume of the gas. The conductivity σ with which we are concerned is the transverse value measured for radio-frequency electric forces of angular frequency p, applied at right angles to an imposed magnetic field H_0.

N, e
m
σ
p
H_0

The symbols retaining the same significance, it is easy to show, from a comparison of the equations of Lorentz* and the equations relating to the propagation of electromagnetic waves through a conducting dielectric, that we can regard the ionized medium as possessing a conductivity, transverse to the magnetic field, given by

$$\sigma=\frac{p}{4\pi}\frac{\beta\,(\alpha^2+\beta^2+\gamma^2)}{(\alpha^2-\beta^2-\gamma^2)^2+4\alpha^2\beta^2} \qquad \text{......(1)},$$

where

$$\alpha=-\frac{mp^2}{4\pi Ne^2} \qquad \text{......(2)}, \qquad \alpha$$

$$\beta=\frac{mp\nu}{4\pi Ne^2} \qquad \text{......(3)}, \qquad \beta$$

$$\gamma=\frac{pH_0}{4\pi Nec}=\frac{mpp_H}{4\pi Ne^2} \qquad \text{......(4)}, \qquad \gamma$$

where

$$p_H=\frac{H_0e}{mc} \qquad \text{......(5)}, \qquad p_H$$

and ν is the frequency of collisions of an electron with the surrounding neutral molecules, while c is the velocity of light.

ν
c

We have here assumed, following Appleton and Chapman$^{(4)}$, that the frictional force per unit velocity, suffered by a vibrating electron, is equal to $m\nu$. We have also, following Darwin$^{(5)}$, omitted the usual Lorentz polarization term.

(*a*) *Case of zero magnetic field.* If there is no imposed magnetic field, equation (1) reduces to

$$\sigma=\frac{p}{4\pi}\cdot\frac{\beta}{\alpha^2+\beta^2} \qquad \text{......(6)},$$

or, on substitution from equations (2) and (3) for α and β

$$\sigma=\frac{Ne^2}{m}\cdot\frac{\nu}{\nu^2+p^2} \qquad \text{......(7)}.$$

This expression reaches a maximum for constant N and p, when $\nu=p$. If, therefore, we find the pressure at which σ is a maximum for constant N and p, we can find ν for that pressure. This is the basis of the method of measuring ν used by Appleton and Chapman.

In discussions of the absorption suffered by radio waves in travelling through the ionosphere it has been usual to assume that since, for a given wave-frequency, the attenuation factor is proportional to σ, the absorption suffered by the waves for given N is a maximum at an atmospheric level for which $\nu=p$. There is now abundant evidence, however, that the propagation of wireless waves is very materially influenced by the earth's magnetic field, since the effective electrical carriers are known to be of electronic mass. It is, therefore, important to find whether this level of maximum absorption is altered by the presence of such a field.

---

* Lorentz, *Theory of Electrons*, equation (239), p. 160. In deriving equation (1) we have neglected the additional electric force which Lorentz includes as arising from the polarization at right angles to the applied electric forces. The significance of this omission is discussed later.

(*b*) *Case of imposed magnetic field.* Equation (1) for the high-frequency conductivity transverse to the magnetic field may be written

$$\sigma=\frac{p}{4\pi}\frac{\beta(\alpha^2+\beta^2+\gamma^2)}{(\alpha^2+\beta^2+\gamma^2)^2-4\alpha^2\gamma^2} \qquad \text{...(8).}$$

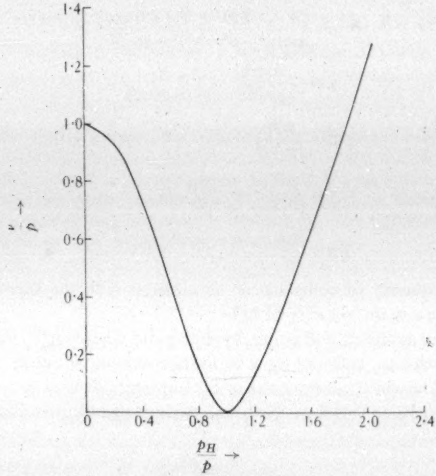


Figure 1.

Substituting from equations (2), (3) and (4) for α, β, γ, we have

$$\sigma=\frac{Ne^2}{m}\frac{\nu(p^2+\nu^2+p_H{}^2)}{(p^2+\nu^2+p_H{}^2)^2-4p^2p_H{}^2} \qquad \text{......(9).}$$

From a practical point of view we are interested in the value of ν for which this expression becomes a maximum, the values of p and p_H being given. Unfortunately it is not possible to obtain this value of ν in a tractable algebraic expression, and we have therefore been compelled to use graphical methods. In figure 1 are given the results of this analysis where the value of ν/p for maximum conductivity is plotted as a function of p_H/p.

This curve shows the very marked effect of p_H in determining the value of ν which produces maximum conductivity. The figure calls for two further comments. It will be seen that when $p_H=0$, so that there is no applied magnetic field, the value of $\nu = p$ for maximum conductivity as was shown above. It should also be noted that according to equation (9), from which the figure was derived, the value of ν for which maximum conductivity is attained when $p_H/p = 1$, is equal to zero. Since, however, a value of ν equal to zero cannot be attained in a tube of finite size, it follows that we must expect divergences between the theory and the experimental results when the applied magnetic field has the so-called critical value, i.e. when $H_0 = mcp/e$.

§ 3. THE EXPERIMENTAL ARRANGEMENTS

The method for investigating the variation of conductivity with pressure was the same as that used by Appleton and Chapman, the ionized medium being a low-pressure gas discharge. The discharge tube selected was 5 cm. in diameter and 50 cm. in length, cold aluminium disc electrodes 20 cm. apart being used. The applied voltage was supplied from an Evershed-Vignolles motor-generator with a saturated diode in the supply circuit. The tube current and thus the magnitude of the ionization was conveniently controlled by varying the filament current of the diode.

The variation of conductivity was studied by the use of external plate electrodes attached to a Lecher-wire system in which was induced a resonant radio-frequency electromotive force. This arrangement is illustrated in figure 2, where T is a section of the discharge tube, LL are the Lecher wires and EE the external plates. These plates which consisted of copper-foil sheets measured 6 × 4 cm. and were set 6 cm. apart. A pair of Helmholtz coils was mounted coaxially with the tube so that a uniform magnetic field could be imposed on that part of the discharge situated between the external plates.

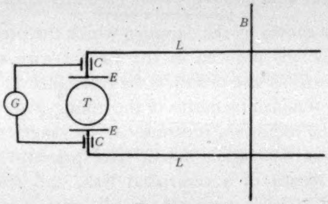


Figure 2.

The tuning-bridge B on the Lecher-wire system consisted of a rectangular copper plate measuring 30 × 20 cm. which could, by means of inlet tubes, be made to slide along the wires with its plane at right angles to their length. The entire space near the free end of the Lecher system was shielded by means of a metal screen and carefully earthed. In this way the effect of the hand-capacity of the observer was so reduced as to be negligible.

The amplitudes of the oscillations induced in the Lecher system from a neighbouring radio-frequency oscillator was, in the case of the longer wave-lengths used (i.e. from 80 to 260 cm.), measured in arbitrary units by means of a Duddell thermogalvanometer coupled to the system by means of two small condensers *CC* consisting of short brass tubes insulated from the Lecher wires by glass sleeves. For the range of wave-lengths studied triode oscillation generators differing in type were used. Some were of the simple regenerative type while others produced internal electronic oscillations of the Barkhausen-Kurz type.

It was however considered desirable to test the method with extremely high radio frequencies, corresponding to wave-lengths of from 15 to 40 cm., and for that purpose a G.E.C. split-anode magnetron oscillator* was used and found to work excellently. The Duddell thermogalvanometer was, however, not found sufficiently sensitive for use with ultra-short waves. For these extremely high frequencies we therefore used a thermocouple made by the Cambridge Scientific Instrument Company and working with a sensitive mirror galvanometer.

In the part of the discharge opposite to the external plates there was fitted a probe electrode for the measurement of electron-concentration by the method described by Langmuir and Mott-Smith(6). By means of this we were able to verify that when the pressure of the air in the tube was varied through the required range, the value of N, the number of electrons per cm^3, did not vary so long as the tube current was maintained by the diode at a constant value.

§ 4. EXPERIMENTAL PROCEDURE

The conductivity is shown by the damping which the presence of the discharge produces in the oscillations induced in the Lecher-wire system. To study the relation between conductivity and pressure, the resonance deflection of the thermogalvanometer was first obtained by means of the tuning-plate bridge. The discharge was then started for the minimum pressure, in the range to be investigated, and the resonance deflection was again noted. The pressure was then allowed to increase in steps by means of a controlled leak, and resonance galvanometer deflections were noted at each pressure. In all cases it was found that at one particular pressure the amplitude of the oscillations was a minimum. This indicated that the conductivity in the discharge was a maximum. Such measurements were made both without an imposed magnetic field and with an imposed field, the magnitude of which was varied.

§ 5. EXPERIMENTAL RESULTS

In order to test the apparatus, measurements were made of the pressure which produced maximum conductivity in the absence of an imposed magnetic field. The range of frequencies used was a little larger than that used by Appleton and Chapman.

---

* We are extremely indebted to the Director of the General Electric Company's Research Laboratories for his kind assistance in this matter and to Mr E. C. S. Megaw of the same laboratories for helpful advice on many points.

In figure 3 are shown some of the readings obtained. Here the resonance amplitude of the Lecher-wire system is shown as a function of the pressure in the discharge tube for three different wave-lengths. In each case it will be seen that the critical pressure is unaltered by a variation of N, brought about by a variation of tube current, but that the critical pressure for maximum conductivity varies with the wave-length.

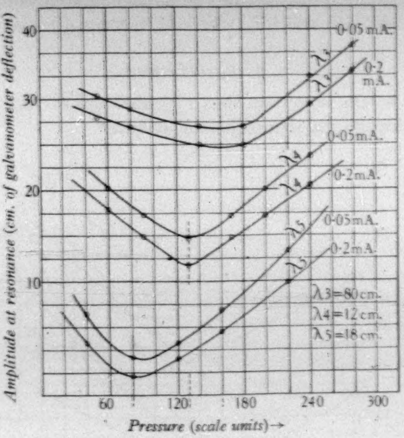


Figure 3.

In the subjoined table are tabulated the results obtained for the different wave-lengths and the values of ν calculated from the relation $\nu = p$; see above, and particularly equation (7).

Table

Wave-lengths (cm.)	Frequency (c./sec. × 10^9)	Pressure X for maximum conductivity (mm. of mercury)	ν (sec.$^{-1}$ × 10^9)	$\frac{\nu}{X}$ (c.g.s.u. × 10^{10})
16	1·88	0·40	11·8	2·95
20	1·50	0·32	9·43	2·95
28	1·07	0·231	6·73	2·91
34	0·882	0·190	5·55	2·92
80	0·375	0·080	2·36	2·95
124	0·242	0·052	1·52	2·92
180	0·166	0·038	1·05	2·76
244	0·123	0·027	0·77	2·85
260	0·115	0·025	0·72	2·88

The constancy of the figures in the fifth column shows, as was expected, that the value of ν is proportional to the pressure.

In figure 4 are shown some of the readings obtained for various values of the imposed magnetic field, the frequency of the applied electromotive forces being kept constant and corresponding to a wave-length of 80 cm. The necessity for keeping the discharge conditions constant during the run rendered it impossible to make measurements at a large number of values of the imposed field. The values chosen were $\frac{1}{2}H_0$, H_0, $\frac{3}{2}H_0$ and $2H_0$, where H_0 was equal to the critical field given by the equation

$$p=p_H=H_0e/mc.$$

H_0 for the case of 80-cm. waves was thus 133 gauss and corresponded to a current of 4·43 amperes through the Helmholtz-coil system.*

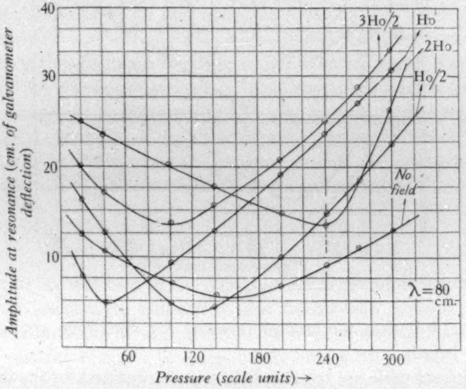


Figure 4.

It will be seen that the imposition of the field had a marked effect on the pressure at which the maximum conductivity was obtained. This is further illustrated in figure 5 where the critical pressure is exhibited as a function of the imposed field.

Now since ν is proportional to pressure we can compare the experimental curve in figure 5 with the theoretical curve in figure 1. Apart from certain points which are discussed later, it will be seen that there is a striking similarity between the two curves. We can, therefore, conclude that both theory and experiment indicate that the critical pressure at which the conductivity of an ionized medium reaches its maximum value is very markedly altered by the imposition of a magnetic field in a direction at right angles to the applied alternating electric forces.

It will be noticed that the experimental results do not agree with the theory for the case in which the magnetic field attains the critical value. This disagreement might be partially explained by the fact, previously mentioned, that under our experimental conditions a zero value of ν is never attained. But we do not believe that this is the only possible reason for the disagreement. It is also probable that, owing to the charges on the walls of the glass tube and the influence of positive-ion sheaths, somewhat complex polarization forces are introduced. These may be expected to be especially important when the quasi-resonance condition is reached and the electron amplitudes are large. We have attempted a quantitative treatment, taking into account the polarization effects of free charges appearing at the boundaries of a dielectric slab of limited extent, and found that in such a case the maximum conductivity for critical applied magnetic field would be expected at small but not zero values of ν; but as we are not sure that such theoretical premises accurately correspond with the experimental conditions we have not pursued the matter further.

---

* Exactly comparable results, not given in the paper, were obtained for a wave-length of 260 cm.

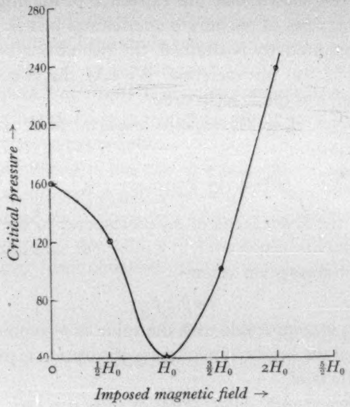


Figure 5.

§ 6. DISCUSSION OF RESULTS

The results summarized above have an important bearing on the current theories of the propagation of wireless waves in the ionosphere. The reflection of such waves by the upper atmospheric electricity is usually accompanied by a certain amount of absorption, and this absorption is attributed to the effects of the collision of electrons with the neutral air molecules. It has more than once been pointed out that the amount of absorption depends very largely on the amount of ionization at a certain critical level in the atmosphere, the height of which depends on the frequency of the waves in question. Thus, for example, if there is no imposed magnetic field the spatial attenuation coefficient κ is given by κ

$$\kappa=\frac{2\pi Ne^2}{c\mu m}\left(\frac{\nu}{\nu^2+p^2}\right) \qquad \text{......(10),}$$

μ

where μ is the refractive index of the medium for waves of angular frequency p. Now the quantity in brackets on the right-hand side reaches a maximum value when $\nu = p$, so that if N/μ is not varying markedly with height there is a maximum of absorption at the particular level in the ionosphere where this condition is attained$^{(7)}$.

It may, however, be shown that the existence of an imposed magnetic field alters the height of this level of maximum attenuation just as it alters the pressure at which maximum conductivity is attained, the two effects being merely different aspects of the same physical phenomenon. We take the two cases of propagation along and at right angles to the field in order.

Case (a): propagation along the imposed magnetic field. Here we have corresponding to equation (10)

$$\kappa = \frac{2\pi N e^2}{c\mu m}\left(\frac{\nu}{\nu^2 + (p \pm p_H)^2}\right) \qquad \text{......(11)},$$

where the upper and the lower signs of p_H correspond to the ordinary and extraordinary wave-components respectively. Under the conditions discussed above there is a maximum of absorption when

$$\nu = p \pm p_H \qquad \text{......(12)}.$$

Now as the value of p_H is comparable with the value of p commonly used in wireless experiments we see that the height of maximum absorption is profoundly influenced by the earth's magnetic field.

Case (b): propagation at right angles to the imposed magnetic field. Here we must distinguish between the case in which the electric vector is parallel and that in which it is at right angles to the imposed magnetic field. In the former case, of the ordinary wave, there is no influence of the magnetic field on the wave-propagation and the equation (10) applies. In this case, therefore, the maximum absorption occurs when $\nu = p$. In the case of the extraordinary wave, where the electric vector is at right angles to the magnetic field, we have, to the degree of approximation we have previously used in this paper,

$$\kappa = \frac{2\pi N e^2}{c\mu m}\,\frac{\nu\,(\nu^2 + p^2 + p_H^2)}{(\nu^2 + p^2 + p_H^2)^2 - 4p^2 p_H^2} \qquad \text{......(13)}.$$

Here the relation between the value of ν for maximum absorption and p_H is similar to that which has been made the subject of experiment in this paper, being shown theoretically in figure 1 and experimentally in figure 5. From these diagrams it may be deduced that for a large range of values of p_H/p the value of ν/p for maximum absorption differs markedly from the value of unity which it assumes under the condition of zero magnetic field.

A further word is necessary to emphasize the difference between the conditions which obtain in the transmission of electric waves and those we have realized in the laboratory. For the propagation of wireless extraordinary waves at right angles to the imposed magnetic field the medium behaves as if the conductivity σ were given by

$$\sigma = \frac{p}{4\pi}\,\frac{\beta\,(1 + 2\alpha + \alpha^2 + \beta^2 + \gamma^2)}{(\alpha + \alpha^2 - \beta^2 - \gamma^2)^2 + \beta^2\,(1 + 2\alpha)^2} \qquad \text{......(1}a\text{)},$$

which reduces to equation (1) when the ionization is small and $|\alpha| \gg 1$. The difference between equations (1 *a*) and (1) is that the former takes account of the electric forces which arise in the direction of propagation and which are brought about by the longitudinal nature of the electronic motion. With a medium of size small compared with a wave-length, such as we have used in our laboratory experiments, we have neglected this effect since in any case it must be small in comparison with the local polarization forces arising from the fact that the medium is limited in extent. But as Lorentz has shown it should be taken into account in electric wave-propagation in an infinite medium and we believe that Pedersen, in his book *The Propagation of Radio Waves*, is unjustified in adopting equation (1) and specifically rejecting equation (1 *a*) as the appropriate expression for the effective conductivity in radio wave-propagation.

An allied topic of interest is the case of direct-current conductivities along and at right angles to the imposed magnetic field. The relevant formulae in this case were given by Townsend many years ago. They are easily derived from our formulae by putting p, the angular frequency, equal to zero. Thus for the conductivity σ_L along the field we have, from equation (7), σ_L

$$\sigma_L = \frac{Ne^2}{m\nu} \qquad \text{......(14).}$$

Correspondingly for the conductivity σ_T transverse to the magnetic field we have from equation (9), σ_T

$$\sigma_T = \frac{Ne^2}{m} \frac{\nu}{\nu^2 + p_H{}^2} \qquad \text{......(15).}$$

Equation (15) is of particular interest, for it indicates that, in higher magnetic latitudes, horizontal direct-current conductivity of the ionosphere is markedly influenced by the earth's magnetic field. The transverse conductivity per electron (or per ion) reaches a maximum value when $\nu = p_H$. Now since p_H is equal to $8{\cdot}4 \times 10^6$ c./sec. for electrons and $1{\cdot}61 \times 10^2$ for, say, nitrogen molecules, it is clear that there will be a level of maximum transverse conductivity per electron and another, higher, level of maximum transverse conductivity per ion. Now at these two levels of maximum conductivity we have

$$\sigma_T = \frac{Ne^2}{2mp_H} \qquad \text{......(16),}$$

or

$$\frac{\sigma_T}{N} = \frac{ec}{2H_0} \qquad \text{......(17),}$$

a result of great interest, for it indicates that the conductivity per electron or per ion is the same at the two levels of maximum conductivity. We thus see that the direct-current transverse conductivity due to ions can be just as important at a certain level as that due to electrons at a lower level, although in the case of the high-frequency conductivity the effects of the ions may be negligible in comparison with those of the electrons at all levels of the ionosphere.

§ 7. ACKNOWLEDGMENT

The work described here was carried out as part of the programme of the Radio Research Board of the Department of Scientific and Industrial Research.

REFERENCES

(1) Appleton and Chapman. *Proc. phys. Soc.* **44**, 246 (1932); see also van der Pol, *Phil. Mag.* **38**, 352 (1919).

(2) Appleton. *Proc. phys. Soc.* **37**, 22 (1925) and *J. Instn elect. Engrs*, **71**, 642 (1932); see also Nichols and Schelleng, *Bell Syst. tech. J.* **4**, 215 (1925).

(3) Lorentz. *Theory of Electrons*, p. 132 (Leipzig, 1909).

(4) Appleton and Chapman. *Proc. phys. Soc.* **44**, 248 (1932); see also Darwin, *Proc. roy. Soc.* A, **146**, 17 (1934).

(5) Darwin. *Proc. roy. Soc.* A, **146**, 17 (1934).

(6) Langmuir and Mott-Smith. *Gen. elect. Rev.* **27**, 449 (1924).

(7) Heising, R. A. *Proc. Inst. Radio Engrs*, N.Y., **16**, 91 (1928).

551.51.053.5

MODULATION-FREQUENCY-CHANGE TECHNIQUE FOR IONOSPHERIC MEASUREMENTS

By GEOFFREY BUILDER, Ph.D., F.Inst.P.

AND

A. L. GREEN, Ph.D., F.Inst.P.

P. N. Russell School of Engineering, University of Sydney, Australia

Received May 22, 1935

ABSTRACT. The paper is a critical examination of Appleton's suggestion that the carrier-frequency-change technique should be modified by varying the frequency of modulation instead of the carrier frequency. Definite conclusions, which are set out in § 5 of the paper, have been reached.

§ 1. INTRODUCTION

Four experimental methods of estimating the effective heights of the ionized regions are known, namely (*a*) the frequency-change device of Appleton and Barnett(2); (*b*) the group-retardation scheme of Breit and Tuve(3); (*c*) the angle-of-incidence method of Appleton and Barnett(4); (*d*) the moving-transmitter arrangement of Mirick and Hentschel(5). It has been shown by Appleton(1) that the first two measure the equivalent path-difference between the ground and sky rays, while the others enable an estimate to be made of the angle of incidence of the sky ray at the ionosphere.

Appleton* has suggested, as especially suitable for ultra-short waves, a modification of the frequency-change method in which the amplitude of the transmitted wave is modulated at a variable audio-frequency but, so far as we are aware, no details have yet been given of the technique of ionospheric measurements made by this new method. In addition, it is not immediately obvious that the carrier-frequency-change and the modulation-frequency-change experiments should measure the same quantity.

---

* See reference (6). The preliminary experiments were made in 1929 with positive results at the Peterborough Radio Research Station by Prof. Appleton and one of us (A. L. G.) by means of transmissions from the B.B.C. stations 2LO and 5SW at frequencies of 0·83 and 12 Mc./sec. An independent investigation was made in Australia by Dr W. G. Baker at about the same time, but the experiments were discontinued when it was learned that the matter was already receiving attention in England.

Further attempts to develop a satisfactory technique were initiated by Prof. Appleton and one of us (G. B.) in 1931, but the experiments were again delayed in view of the rapid development of the echo work(7).

The present technique is therefore closely related to previous work of both the British and the Australian Radio Research Boards, and in particular the new suppressed-carrier arrangement is the basis of much of our recent work on control of fading by special systems of modulation at the transmitting station.

(b) (a)

The object of the present paper, therefore, is firstly to supply this deficiency and to bring the modulation-frequency-change device into line with the four basic experimental methods. Secondly it is desired to put forward a further modification of the modulation-frequency-change arrangement in which the carrier wave is suppressed at the transmitter, only the modulation sidebands being emitted. *In what follows we refer to the suppressed-carrier modification as system* (*b*) *and to Appleton's modulation-frequency-change device as system* (*a*). It will be shown that the former arrangement has some marked advantages over the latter, both in the actual technique of measurement and in the interpretation of results. Recent investigations[(8)] involving the measurement of equivalent path-differences have demonstrated the superiority of system (*b*) in this type of experiment.

§ 2. THEORY OF MODULATION-FREQUENCY-CHANGE EXPERIMENT, SYSTEM (*a*)

E_0, E_s r θ

Before proceeding to the investigation of system (*a*), it will be convenient, for the purpose of comparison, to summarize very briefly the salient features of the carrier-frequency-change experiment. If E_0, E_s are the vertical components of the electric intensities of the ground and sky rays of radio-angular frequency r, and θ is the phase-difference between them, then the signal at the receiver is

$$E_0 \sin rt + 2E_s \sin(rt+\theta) \qquad \text{......(1)},$$

where

$$\theta = -\frac{rD}{c} \qquad \text{......(2)},$$

D, c

D is the optical path-difference between the ground and sky rays, and c is the velocity of electromagnetic radiation in free space.

n

When a frequency-change of amount Δr is made at the transmitter, the received signal passes successively through a number Δn of interference fringes, determined by

$$\Delta n = -\frac{\Delta r}{2\pi}\cdot\frac{d\theta}{dr} = \frac{\Delta r}{2\pi c}\cdot\left(D + r\cdot\frac{dD}{dr}\right)$$
$$= \frac{D'}{c}\cdot\frac{\Delta r}{2\pi} \qquad \text{......(3)},$$

D'

where D' is defined as the equivalent path-difference and is the quantity measured in the experiment.

M, m, A

Further, if the receiver includes a square-law detector and the maximum and minimum deflections of the galvanometer in its anode circuit are denoted respectively by M^2 and m^2, then we define a quantity A which determines the amplitude of the fringes to be such that

$$A = \frac{M^2 - m^2}{M^2 + m^2} = \frac{2a}{1+a^2} \qquad \text{......(4)},$$

a

where a is written for $2E_s/E_0$. From equations (3) and (4) it is therefore possible to calculate the equivalent path of the sky ray and the relative intensities of the ground and sky rays.

In the modulation-frequency-change system (*a*), three waves are simultaneously emitted and the signal at the receiver due to the ground rays is

$$E_0 \left[\sin rt + \frac{Q}{2} \{\sin (r+q)\, t + \sin (r-q)\, t\} \right]$$

which

$$= E_0 (1 + Q \cos qt) \sin rt \qquad \text{......(5)},$$

where q is the angular frequency of modulation of depth Q. The sky-ray signals at the receiver are q, Q

$$2E_s \left[\sin (rt+\theta) + \frac{Q}{2} [\sin \{(r+q)\, t + \theta'\} + \sin \{(r-q)\, t + \theta''\}] \right] \qquad \text{......(6)},$$

where θ', χ

$$\theta' = \theta + q \frac{d\theta}{dr} = \theta + \chi,$$

$$\theta'' = \theta - q \frac{d\theta}{dr} = \theta - \chi \qquad \text{......(7)},$$

θ''

and, from equation (2),

$$\chi = q \frac{d\theta}{dr} = -q \frac{D'}{c} \qquad \text{......(8)}.$$

Hence, the sky-ray signals may also be written as

$$2E_s \{1 + Q \cos (qt + \chi)\} \sin (rt + \theta) \qquad \text{......(9)},$$

which is to be compared with equations (1) and (5) above.

After square-law detection of the combined ground and sky-ray signals, the d.-c. components in the galvanometer are proportional to

$$(E_0^2 + 4E_s^2)(2 + Q^2) + 4E_0 E_s (2 + Q^2 \cos \chi) \cos \theta \qquad \text{......(10)},$$

in which it should be noted that θ is invariant with respect to q.

As the frequency of modulation is changed, $\cos \chi$ varies between the limits ± 1 and the number of fringes Δn produced by a change Δq of modulation angular frequency is given by

$$\Delta n = -\frac{\Delta q}{2\pi} \cdot \frac{d\chi}{dq} = \frac{D'}{c} \cdot \frac{\Delta q}{2\pi} \qquad \text{......(11)}.$$

On comparing this result with equation (3), it will be found that both the carrier-frequency-change and the modulation-frequency-change experiments measure the equivalent path-difference D'. However, in the latter case the amplitude of the observed fringes is A_1 where A_1

$$A_1 = \frac{2a\, Q^2 \cos \theta}{(1 + a^2)(2 + Q^2) + 4a \cos \theta} \qquad \text{......(12)},$$

from which it follows that A_1 depends not only on the sky-ray intensity, but also on the depth of modulation and on the instantaneous phases of the ground and sky rays. Even under the most favourable experimental conditions it is found that the fringe-amplitude is very much less than that to be observed with the carrier-frequency-change technique and, when the phase-difference between the ground and sky rays is such that $\cos \theta = 0$, the fringes are completely eliminated.

In practice it is desirable to make use of an artifice* which will facilitate the delineation of the fringes. The required effect has been achieved by passing the audio-frequency signal components from the anode circuit of the detector to a second square-law rectifier to which the galvanometer is connected.

From equations (5) and (9) it may be shown that the audio-frequency current components in the first detector are proportional to

$$E_0^2\,(4Q\cos qt+Q^2\cos 2qt)+4E_s^2\,\{4Q\cos(qt+\chi)+Q^2\cos 2\,(qt+\chi)\}$$
$$+4E_0E_s\cos\theta\left\{4Q\cos\frac{\chi}{2}\cos\left(qt+\frac{\chi}{2}\right)+Q^2\cos(2qt+\chi)\right\}\quad\ldots\ldots(13),$$

and, after rectification of this signal in the second detector, the d.-c. components passing through the galvanometer are proportional to

$$Q^2\,[1+a^4+2a^2\cos\chi+2a\cos\theta\,(1+\cos\chi)\,(a\cos\theta+1+a^2)]$$
$$+\frac{Q^4}{16}[1+a^4+2a^2\cos 2\chi+4a\cos\theta\,\{a\cos\theta+(1+a^2)\cos\chi\}]\quad\ldots\ldots(14).$$

It will be observed that the harmonic distortion in the first detector, of frequency $2q$, has given rise to a term containing $\cos 2\chi$ which will produce spurious secondary fringes on the records. However, the relative amplitudes of the subsidiary to the primary fringes are proportional to $Q^2/16$ and, in practice, it is found that the effect is negligible if the depth of modulation at the transmitter is limited to values of the order of 30 per cent.

With this condition fulfilled, the number of fringes is determined solely by variations of $\cos\chi$ and is in fact given by equation (11), as for the first detector. On the other hand, the fringe-amplitude is now A_2 where

A_2

$$A_2=\frac{4a\,(a+\cos\theta)\,(1+a\cos\theta)}{(1-a^2)^2+(1+a^2+2a\cos\theta)^2}\quad\ldots\ldots(15),$$

and, except for the two special cases in which the intensity and phase of the sky ray are such that $\cos\theta=-a$ or $-1/a$, the fringes are easily distinguished. In fact, when $\cos\theta=+1$, equation (15) reduces to $A_2=2a/(1+a^2)$, showing that in this case the fringe-amplitude is identical with that which would have been observed with carrier-frequency-change technique. It should also be noticed that A_2 is independent of the depth of modulation at the transmitter, whereas the fringe-amplitude in the first detector is directly proportional to Q^2.

Of the two special cases cited above in which the fringe-amplitude falls to zero, only the first is of practical importance. The second condition clearly can only be satisfied when the sky-ray signal is greater than that of the ground ray, and this is usually avoided in frequency-change experiments.

An alternative equation which has been found somewhat more convenient than equation (15) for calculation, is derived in the following way. We extract the

---

[*] In the preliminary experiments carried out by Prof. Appleton at the Peterborough Radio Research Station the same principle was made use of.

F

square roots M and m of the maximum and minimum galvanometer-deflections and define a quantity F as a measure* of the signal fluctuations, thus

$$F=\frac{M-m}{M+m}=\frac{a\,(a+\cos\theta)}{1+a\cos\theta} \qquad \text{......(16).}$$

This equation should be used with care if it is suspected that the sky-ray signal is greater than that of the ground ray. Signal-fluctuations are greatest when $\cos\theta=+1$ and when $a=1$; there are no signal-variations when $a=0$ or $-\cos\theta$, in accordance with equation (15) above.

§ 3. THEORY OF MODULATION-FREQUENCY-CHANGE EXPERIMENT, SYSTEM (*b*)

Measurements carried out with system (*a*) have disclosed a number of difficulties in the technique of the receiving-apparatus, of which the most serious is the extreme susceptibility to interference by atmospherics and by adjacent continuous-wave transmissions. Both types of interference were found to be considerably enhanced by the use of the second detector in the receiver, and the suggestion to suppress the carrier at the transmitter was made as a result of a search for a method which would obviate the necessity of employing successive rectifiers. It was realized that the amplitude of the fringes after one stage of detection would then be sufficiently great for recording purposes and, indeed, subsequent work has shown that when system (*b*) is used no advantage is to be expected from the addition of the second detector. Since it is not necessary to resupply the carrier at the receiver it is possible to use the same apparatus for modulation-frequency-change experiments as is required in carrier-frequency-change technique.

In system (*b*) only the modulation sidebands are emitted and the signal at the receiver due to the ground rays is proportional to

$$\frac{E_0}{2}\{\sin(r+q)\,t+\sin(r-q)\,t\}=E_0\cos qt\sin rt \qquad \text{......(17),}$$

while the corresponding sky-ray signal is

$$E_s\,[\sin\{(r+q)\,t+(\theta+\chi)\}+\sin\{(r-q)\,t+(\theta-\chi)\}]$$

which

$$=2E_s\cos(qt+\chi)\sin(rt+\theta) \qquad \text{......(18).}$$

This is to be compared with equations (1), (5) and (9) above.

In the first detector the d.-c. components which actuate the galvanometer are proportional to

$$E_0^2+4E_s^2+4E_0E_s\cos\theta\cos\chi \qquad \text{......(19).}$$

Thus the number of fringes produced by a change in the modulation frequency is determined by variations of $\cos\chi$, and the equivalent path-difference between the ground and sky rays may be computed, as in the case of system (*a*), from equation (11), thus

$$\Delta n=\frac{\Delta q}{2\pi}.\frac{D'}{c}.$$

---

* In the carrier-frequency-change experiments it is often convenient to use F in preference to A. For comparison with equation (4) above, $F=a$ under the same conditions.

It is important to notice, however, that the amplitude of the fringes is now given very simply by the equation

$$A=\frac{2a\cos\theta}{1+a^2} \qquad \text{......(20).}$$

Hence when the instantaneous phase between the ground and sky rays is such that $\cos\theta=1$, the amplitude of the fringes is as great as would have been obtained with carrier-frequency-change technique. At other times there is a definite reduction in fringe-amplitude and, in particular, no fringes are observable when $\cos\theta=0$.

§ 4. EXPERIMENTAL RESULTS

A number of comparisons have been made of ionospheric heights by the carrier-frequency-change and the modulation-frequency-change techniques at a mean frequency of 1·45 Mc./sec. The receiver was located at Liverpool, New South Wales, distant 25 kilometres from the emitter at Sydney University. In all cases substantial agreement has been found between the two methods, subject to experimental errors in the measurement of the amounts of the frequency-change in the two techniques.

For example, on October 1, 1934, when the carrier-frequency-change was 6300 c./sec. and the modulation frequency was varied between the limits 900 and 6300 c./sec., pairs of records were obtained at short intervals starting from 10 p.m. E.S.T., and the equivalent heights computed were as shown in the table.

Time, p.m. (E.S.T.)	10.00	10.10	10.20	10.25	10.30	10.35	10.40
Heights by carrier-frequency-change	107	107	258	256	260	255	262
Heights by modulation-frequency-change	109	109	262	262	251	253	260

It will be observed that a jump occurred from the *E* to the *F* region at 10.15 p.m. and that the two series of measurements were in good agreement for both regions.

On another occasion, September 28, 1934, measurements were made during the sunset period during which there was a steady increase of height of the *E* region, followed by a jump to the *F* region at 6 p.m. Sunset on this day was at 5·55 p.m. The results of the measurements by the two methods are shown in figure 1, from which an idea may be gained of the order of agreement to be expected between the modulation-frequency-change and the carrier-frequency-change techniques.

A careful check has also been made of the manner in which the amplitude of the fringes in the modulation-frequency-change experiment depends both on the intensity and on the instantaneous phase relationships of the ground and sky rays. Figures 2 and 3 show the results, for a typical set of experimental conditions, which would be expected to follow from the theory previously given. In both cases the experimental conditions correspond to reflection of sky rays from the *E* region at nearly vertical incidence, and it has been assumed that $D'=225$ km. and $a=0{\cdot}5$. Points in the curves have been computed for a change of modulation frequency from 500 to 2000 c./sec., but the extensions required for greater frequency-changes will be easily visualized.

Figure 2, for modulation-frequency-change system (*a*), has been computed by substituting a range of values of θ in expression (15). The critical phase angle between the ground and sky rays which determines the elimination of fringes is given by $\cos\theta=-a$, and is 120° for the experimental conditions chosen. When the ground- and sky-ray signals corresponding to the carrier are in phase, $\theta=0°$ and

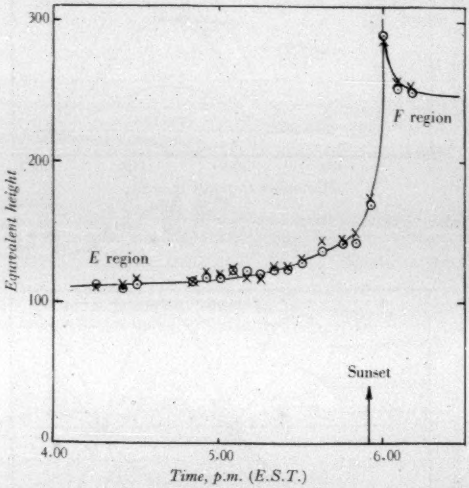


Figure 1. Comparison of ionospheric heights measured by modulation-frequency-change of 5370 c./sec. ⊙ ⊙, and carrier-frequency-change of 6000 c./sec. × ×.

the fringe-amplitude is a maximum and is as great as would have been observed in a carrier-frequency-change experiment. For increasing values of θ the fringe-amplitude steadily decreases through zero until finally there is a reversal in sense. However, in the special case when $a=1$ there is no reversal, but both the fringe-amplitude and the mean signal attain zero simultaneously. It should be noted that the beginning of the fringe occurs in every case at the same point of a cycle, independently of the instantaneous value of θ. This result is in striking contrast with the carrier-frequency-change case in which the commencement of the artificial fading cycle is determined by the momentary state of the natural fading.

Figure 3, for modulation-frequency-change system (*b*), is calculated from equation (20). In this case the fringes always have zero amplitude when $\theta=90°$, independently of the intensity of the sky ray. The fringe-amplitude is a maximum and equal to that observed in a carrier-frequency-change experiment both when $\theta=0°$ and $180°$. Apart from the reversal in sense as θ passes through 90°, the commencement of each fringe is at the same point in the cycle, independently of the phase of ground and sky rays.

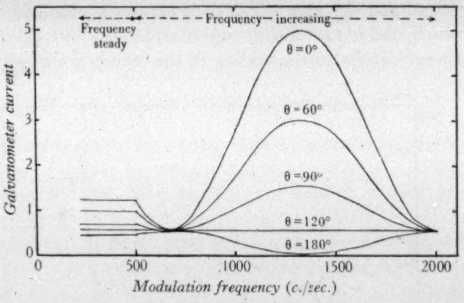


Figure 2. Theoretical variation of fringe-amplitude in the modulation-frequency-change method with normal modulation, compared with the radio-frequency phase angle θ between the ground and sky rays.

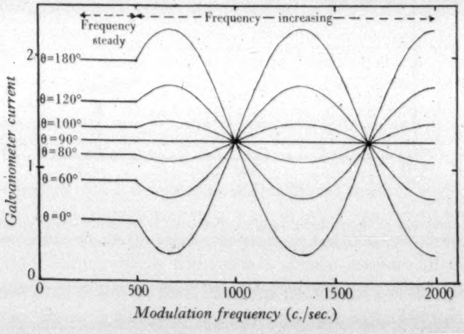


Figure 3. Theoretical variation of fringe-amplitude in the modulation-frequency-change method with suppressed carrier.

Drawings from typical records chosen to illustrate these points are shown in figures 4, 5 and 6. In figure 4 there are three sets of fringes from the same record of the anode current in the first detector, for system (*a*) and a depth of modulation of about 30 per cent at the transmitter. It is seen that the fringe-amplitude was always small and, at the point marked *B*, very nearly zero; at that time the phase angle between ground and sky rays was 90°. At *A* and *C*, when the state of the natural fading showed that the two rays were respectively in phase and in antiphase, the fringe-amplitude was at a maximum. The interesting sense-reversal of the fringes, which is predictable from equation (12), is apparent on comparison of *A* and *C*. For this record the modulation-frequency-change was 5·5 kc./sec., so that the 3·8 fringes indicated an equivalent path-difference of 207 km. A set of fringes obtained under comparable conditions but with the carrier-frequency-change system are shown at *D*, and the comparison shows clearly the need for the second detector in the modulation-frequency-change experiments when system (*a*) is used.

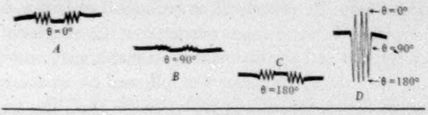


Figure 4. Experimental comparison of interference fringes obtained with carrier-frequency-change and modulation-frequency-change with normal modulation. *A*, *B*, *C*, first detector in system (*a*); *D*, carrier-frequency-change system.

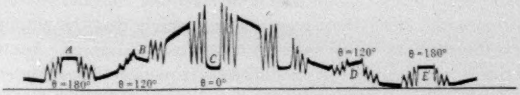


Figure 5. Experimental verification of variation of fringe-amplitude with radio-frequency phase angle θ between ground and sky waves. Modulation-frequency-change with normal modulation, was employed and the fringes were recorded in the output of the second detector.

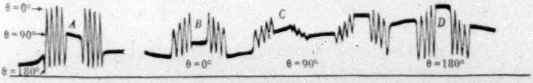


Figure 6. Comparison of carrier-frequency-change method. *A*, with modulation-frequency-change; *B*, *C*, *D*, suppressed-carrier method.

Figure 5 is for the second detector in tests made with modulation-frequency-change system (*a*). Six complete sets of fringes are shown, each comprising an increase and a decrease of modulation frequency of amount 5·5 kc./sec. At *C*, in the middle of the record, the ground and sky rays were in phase and the fringe-amplitude was a maximum. At *A* and *E*, the beginning and end of the record, the ground and sky rays were in antiphase and the fringes had undergone the sense-reversal predicted in figure 2 from equation (15). It is known* that the modulation frequency was steady and at its minimum value at the points marked *A*, *B*, *C*, *D*, *E*, and it may be seen from the record that the sense of the fringes was such that the received signal at *C* at first increased as the modulation frequency began to increase, whereas at A and E frequency-increases were accompanied by decreases of signal. During the time that this record was being taken, natural fading was occurring at a rate somewhat too fast to enable the reduction of fringe-amplitude to zero to be shown throughout a complete set of fringes. However, the transition, corresponding to $\cos\theta = -a$, is clearly observed at B and it seems likely that the phase-difference between ground and sky rays was about 120° at this time, in agreement with equation (15).

---

* The apparatus which produces the frequency-changes is so arranged that there is a short pause of about 1 sec. duration following the decrease of frequency and a longer pause of about 2 sec. after the frequency-increase.

Figure 6 is for the first detector in modulation-frequency-change system (*b*). At A are shown, for comparison, a set of fringes taken with the carrier-frequency-change arrangement 1 min. previously. The reduction of fringe-amplitude to zero and the reversal in sense of the fringes occurred at C, at which time the phase-difference between ground and sky rays was presumably 90°, according to equation (20). At B, when $\theta = 0°$, the short pause was followed by an increase of frequency and by an increase of signal, whereas at D, when $\theta = 180°$, the fringe immediately following the short pause began with a decrease of signal-intensity.

Many other records have been obtained with both systems (*a*) and (*b*), and agreement has always been found with the theory given here. It is therefore clear that modulation-frequency-change suffers at least one considerable disadvantage when compared with carrier-frequency-change, namely that the amplitude of the fringes is, on the average, much less with the former technique. Another experimentally verified peculiarity of the modulation-frequency-change system, that the phase at the beginning of a set of fringes is substantially independent of the instantaneous state of the natural fading, has an important bearing on measurements of polarization and will be discussed later.

§ 5. DISCUSSION AND CONCLUSIONS

Although this paper is mainly concerned with the technique of ionospheric measurements by means of modulation-frequency-change systems it does not appear to be necessary to describe the apparatus in great detail. We are satisfied that system (*b*) is the more satisfactory of the two modulation-frequency-change arrangements, so that if this system is used it is unnecessary to make any alterations to receiving-apparatus which has already been set up for carrier-frequency-change work. As to the transmitter, it is impossible to use the simple arrangement comprising a self-excited oscillator which has hitherto proved adequate for most carrier-frequency-change experiments. However, the substitution of a low-power master-oscillator and amplifying stages has been found to be an improvement in transmitter technique for carrier-frequency-change work, so that when once this replacement has been made it is an easy matter to arrange for carrier-suppression. The system which we have found useful is shown in figure 7. The output from a beat-frequency audio-frequency oscillator is applied in push-pull to the grids of two valves through a transformer at A, whereas the carrier is supplied at B to the two valves in phase; radio-frequency chokes and resistances serve to isolate radio- and audio-frequency circuits. When the valves are carefully matched the output at C consists only of the modulation sidebands, since the carrier is balanced out and the output tuned circuit presents negligible impedance to currents of audio frequency.

It has already been pointed out that the phases of the beginnings and ends of the fringes in a modulation-frequency-change experiment are, with the exception of the phase-reversal, independent of the momentary phase relations of the ground and sky rays. On the other hand, a knowledge of these phases in carrier-frequency-change work has made possible the measurement of the polarization(9, 10) of the sky ray, of lateral deviation(11, 12), and of changes in optical path(13). The following analysis shows that the modulation-frequency-change system is not suitable for these experiments and, in fact, it is easy to see that its use is very largely limited to measurements of equivalent path.

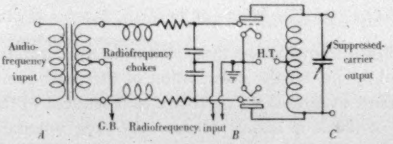


Figure 7. Carrier-suppression unit. The circuit is adapted from carrier telephony practice for use at high radio frequencies.

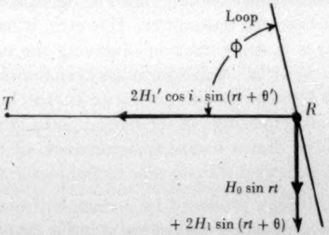

Loop
ϕ
$2H_1' \cos i \,.\, \sin (rt + \theta')$
T
R
$H_0 \sin rt$
$+ 2H_1 \sin (rt + \theta)$

Figure 8. Ground plan of magnetic vectors of ground and sky waves.

Consider the signals received in a rotatable loop, figure 8. It will only be necessary to work out the case for system (*b*), since similar results hold for system (*a*). If the plane of the loop makes an angle ϕ with the plane of propagation, then the signal due to the reception of a ground ray H_0 and an elliptically polarized sky ray H_1, H_1', arriving at an angle of incidence i, is

ϕ

H_0

H_1, H_1', i

$$H_0 \cos \phi \cos qt \sin rt$$
$$+ 2H_1 \cos \phi \cos (qt+\chi) \sin (rt+\theta)$$
$$+ 2H_1' \cos i \sin \phi \cos (qt+\chi) \sin (rt+\theta') \quad \ldots\ldots(21),$$

if we assume that the polarization of sky rays of frequency $(r+q)$ and $(r-q)$ is not very different from that at frequency r. In that case the polarization of the down-

R_c coming ray is determined by the ratio R_c of components, which $=(H_1'/H_1)\cos i$,
ξ and their phase-difference ξ which $=\theta'-\theta$.

The signal in the loop may be rewritten as

$$H_0 \cos\phi \cos qt \sin rt + 2H_s \cos(qt+\chi)\sin(rt+\theta_s) \qquad \text{......(22)},$$

H_s where

$$H_s = H_1 \cos\phi \sqrt{(1+R_c^2 \tan^2\phi + 2R_c \tan\phi \cos\xi)} \qquad \text{......(23)},$$

θ_s and

$$\tan\theta_s = \frac{\sin\theta + R_c \tan\phi \sin\theta'}{\cos\theta + R_c \tan\phi \cos\theta'}.$$

It should be remembered that both H_s and θ_s are invariant with respect to changes in the modulation frequency. The similarity in form of equations (22), (17) and (18) shows that the currents passing through the galvanometer after square-law detection of the signals are proportional to

$$H_0^2 \cos^2\phi + 4H_s^2 + 4H_0 H_s \cos\phi \cos\theta_s \cos\chi \qquad (24),$$

from which it follows immediately that, although the amplitude of the fringes depends on H_s and θ_s, i.e. on the polarization of the downcoming ray, the fringe phases are determined by χ, and are therefore independent of the polarization and of the orientation of the loop aerial. Hence, any experimental method of determining polarization, such as that due to Appleton and Ratcliffe, which depends for its success on recognizing variations in fringe-phases in differently oriented receiving aerials, is inapplicable to the modulation-frequency-change technique.

Similar considerations rule out the use of modulation-frequency-change systems in the measurement of lateral deviation and changes in optical paths of the sky ray. It is perhaps to be emphasized that we have tacitly assumed throughout that the receiving apparatus is sufficiently broadly tuned to be capable of receiving both of the sidebands sent out from the transmitter. However, it has been suggested that there may be advantages in some cases in observing the two signals in separate sharply tuned receivers. With this technique it has been found possible, for example, to emit two waves with frequencies differing by 40 kc./sec. but with known relative phases, and thus to compare changes in the optical paths of the two sky rays.

The conclusions to be drawn from this comparison of the carrier-frequency-change and the modulation-frequency-change techniques are as follows:

(i) The number of fringes produced by a change of modulation frequency is proportional to the equivalent path of the sky ray, as in the case of a carrier-frequency-change.

(ii) The amplitude of the fringes in modulation-frequency-change is very much less than in carrier-frequency-change, unless an artifice such as the two square-law detectors in Appleton's system (*a*), or the device of suppressing the carrier at the transmitter in system (*b*), is used.

(iii) After either of these modifications have been made, the fringe-amplitude is occasionally as great as would have been obtained with the carrier-frequency-change technique. There are certain relations, however, between the carrier phases and amplitudes of the ground and sky rays for which the modulation-frequency-change fringes have zero amplitude, and on the average the carrier-frequency-change fringes always have the greater amplitude.

(iv) In the modulation-frequency-change experiment, the phase at the beginning of a set of fringes is independent of the momentary state of the natural fading. It is also independent of the state of polarization of the sky ray and of the orientation of the receiving loop aerial. Hence, modulation-frequency-change technique is inapplicable to experiments where fringe-phases are important; for example measurements of polarization, lateral deviation, or variations of optical path.

(v) A comparison of modulation-frequency-change systems (*a*) and (*b*) suggests that the latter is the more satisfactory technique for ionospheric measurements. There is little to choose between the two methods in regard to the modifications which would need to be made to a carrier-frequency-change transmitter, but there are marked advantages in system (*b*) at the receiver, both in reception technique and in the interpretation of records.

§ 6. ACKNOWLEDGMENTS

The work described here formed a part of the programme of the Radio Research Board of the Australian Commonwealth Council for Scientific and Industrial Research, to which the authors are indebted for permission to publish their results. They are also indebted to Mr A. H. Mutton and Dr O. O. Pulley for assistance at the transmitter, to Dr D. F. Martyn for criticism of the paper, and to Prof. J. P. V. Madsen of the P. N. Russell School of Engineering of the University of Sydney for much help and advice during the course of the investigation.

REFERENCES

(1) Appleton. *Proc. phys. Soc.* **41**, 43 (1928); **42**, 321 (1930).
(2) Appleton and Barnett. *Nature*, Lond. (March 7, 1925).
(3) Breit and Tuve. *Phys. Rev.* **28**, 554 (1926).
(4) Appleton and Barnett. *Proc. roy. Soc.* A, **109**, 621 (1925).
(5) Mirick and Hentschel. *Proc. Inst. Radio Engrs*, N.Y., **17**, 1034 (1929).
(6) Appleton. *Proc. phys. Soc.* **42**, 323 f.n. (1930).
(7) Appleton and Builder. *Nature*, Lond. (June 27, 1931).
(8) Green and Builder. *Control of Wireless Signal Variations* (in press). It has been found convenient to use substantially the same symbols in the two papers.
(9) Appleton and Ratcliffe. *Proc. roy. Soc.* A, **117**, 576 (1928).
(10) Green, A. L. *Proc. Inst. Radio Engrs*, N.Y., **22**, 324 (1934).
(11) Cherry and Martyn. *Radio Research Board Report*, No. 4. *Bull. Counc. sci. industr. Res. Aust.* No. 63 (1932).
(12) Martyn and Green. *Proc. roy. Soc.* A, **148**, 104 (1935).
(13) Appleton. *Proc. roy. Soc.* A, **126**, 542 (1930).

621.396.621:551.51.053.5 538.56:535.541

A RECEIVER DISCRIMINATING BETWEEN RIGHT AND LEFT-HAND CIRCULARLY POLARIZED WIRELESS WAVES

By O. O. PULLEY, Ph.D., B.Sc., B.E. The Halley Stewart Laboratory, King's College, London

Communicated by Prof. E. V. Appleton, F.R.S., May 14, 1935

ABSTRACT. A brief review of the history of magneto-ionic effects in the ionosphere and the apparatus developed for their investigation is given. Because of the inflexibility of present methods of determining the polarization of received waves, a new scheme has been developed which permits observations to be made over a range of frequencies without readjustment. The principle employed is that of an effectively rotating loop, the rotation being simulated by the modulation in phase quadrature of two loops at right-angles. The theory of operation, the errors and the method of adjustment are investigated. Adjustment can be made with the aid of a test wave of unknown polarization.

Some results obtained with the apparatus are discussed, and it is concluded that the two magneto-ionic components are almost circularly polarized, and that the relative intensities are unpredictable unless the complete $\{P', f\}$ curve appropriate to the time of observation is available. A Fourier-series treatment of the propagation of a pulse of energy of radio frequency in the ionosphere is given in an appendix.

§ 1. HISTORICAL

As soon as an attempt was made to use wireless waves for the purpose of determining position by measurement of the direction of arrival of such waves from a known station, it became evident that this direction, as found from the position of a loop aerial when the received signal was a maximum or a minimum, differed from the expected or true direction, and furthermore, that the difference varied with time. To account for this, it was suggested by Bellini(1) and by T. L. Eckersley(2) that the error is connected with the presence of an abnormally polarized component in the received wave, i.e. a component in which the vector of electric force does not lie in the vertical plane through sender and receiver. To explain the presence of such abnormally polarized waves, the magneto-ionic theory was put forward by Appleton(3), who showed that if the effective agents in the upper atmospheric reflection were of electronic mass, the effect of the earth's magnetic field would be to cause double refraction and rotation of the plane of polarization. Further work has amply demonstrated the correctness of Appleton's theory.

Smith-Rose and Barfield(4) made a systematical study of the effect by comparing the relative signal-amplitudes from various combinations of loop and vertical aerials. Later, Appleton and Ratcliffe (5) obtained direct confirmation of the theory, which was further strengthened by the results obtained by Green (6) working under similar conditions but in the southern hemisphere. The method used was for studying the variation of signal-intensity as measured on three loops, one in the plane of propagation and the other two making angles of $\pm 45^\circ$ with this direction, when the frequency of the transmitted waves was varied. About the same time, Hollingworth (7) made a similar investigation on the longer wave-lengths with apparatus which was slightly simpler but required less acceptable assumptions.

With the adoption of the pulse method of investigation, Appleton and Builder (8) showed the composite nature of the echoes from the upper region, and interpreted the doublets as being due to magneto-ionic splitting. Apparatus was then developed at the Radio Research Board at Slough (9) for demonstrating the polarization of the components of the doublets. The method consists of applying the output, suitably amplified, from two loops at right-angles to pairs of plates in a cathode-ray oscillograph. The resulting Lissajou's figure is an indication of the state of polarization of the received wave, and a simple test determines the sense.

Almost simultaneously, J. A. Ratcliffe and E. L. C. White (10) developed their receiver for circularly polarized waves, in which two loops at right angles are coupled through a mutual inductance. When the receiver is suitably adjusted, a circularly polarized wave having one sense of rotation produces an e.m.f. at the grid of the amplifying valves, while one of opposite sense produces no effect. Reversal of the sense of the mutual coupling permits the reception of the wave polarized in the opposite sense while suppressing the original one. T. L. Eckersley (11) originated a somewhat similar scheme, but detuned the two loops in opposite directions in order to effect the cancellation of one sense of polarization.

The results in all cases were in agreement with Appleton's magneto-ionic theory as far as the actual polarization of the doublets was concerned, though some differences of opinion as to the relative intensities of the two components still exist. (12)

German workers (13) have investigated the problem from the point of view of the Fourier components of the pulse, and so deal with continuous waves of various frequencies rather than with a discrete train of waves of a single frequency. Assuming the magneto-ionic theory to account for a path-difference between the two circularly polarized components of a wave of given frequency, they observe that the plane of polarization of the resultant wave at the receiver varies with frequency. The summation of these component plane-polarized waves, with due allowance for the rotation of the plane of polarization with frequency, results in a complex pulse, usually a doublet. It is shown in appendix C that this doublet is the resultant of two pulses, identical in shape with the emitted pulse, and both circularly polarized, but that when the two pulses overlap the polarization varies throughout the overlapping region, and the shape of the resultant pulse bears only an indirect relation to the shape of the emitted wave since the two components have to be added vectorially. The result is quite independent of frequency dispersion and, incidentally, this factor will not adequately account for the spreading or broadening of echoes of one polarization found experimentally. Both the theoretical and experimental results of Handel and Plendl merely confirm the accepted magneto-ionic theory.

The magneto-ionic theory, then, is well established and is of considerable importance, not only in the transmission of wireless waves but also in several branches of geophysics. But a slight uncertainty with regard to certain points in itself justifies another approach to the problem, and the inflexibility, with respect to frequency-change, of all present schemes makes such an undertaking very desirable.

§ 2. DISCUSSION OF AVAILABLE SCHEMES

All schemes at present in use for the determination of the state of polarization of wireless waves require considerable care in adjustment and this adjustment is necessary for each frequency. In their present state it would not be possible to make observations on more than one frequency in, say, 5 min., and if a range of frequencies sufficient to yield a curve of the variation of height with frequency had to be covered, then conditions at the end of the run would be quite different from those at the beginning. The two-receiver method of the Radio Research Board is very convenient for visual observation, but is unsuitable for recording purposes. It is expensive, since two separate receivers are necessary; furthermore, the constancy of adjustment is likely to be poor, since there are many circuits in which variations can occur.

For recording purposes, the polarization ellipse must be reduced to its elements. These elements might be the magnitude of the major and minor axes, their angle with respect to a reference line, and the sense in which the ellipse is described. Alternatively, the magnitude of the two circular components having opposite senses of rotation and their phase relation might be used. Since it is to be expected that the polarization of the reflected waves will be nearly circular, the latter scheme is preferable; the phase relation is usually omitted for simplicity.

The methods of Ratcliffe and White(10) and of Eckersley(11) can be used for recording the two circular components but, apart from their inflexibility, it is necessary to assume that the testing wave is linearly polarized; as yet, no technique has been developed for adjustment when this wave is, of necessity, elliptically polarized.

All methods are very critical in adjustment, for they depend on the phase of the currents in the loops, which is much more sensitive to incorrect tuning than is the amplitude-adjustment. When a change to another frequency is made, an adjustment other than retuning is necessary—in one case in the amount of coupling, and in the other in the amount of detuning. Finally, in both of these schemes, when it is desired to change from the suppression of one sense of rotation to the suppression of the other, it is necessary to perform a switching operation in a circuit carrying currents of radio frequency, and it is difficult to ensure that circuit conditions shall not be altered in the process.

§ 3. A SUGGESTED SCHEME

If a loop rotates in a clockwise or right-hand direction when viewed from above, then it is obvious that a circularly polarized wave of right-hand sense of rotation will produce in the loop an e.m.f. of frequency less than that of the incident wave, whereas one of left-hand sense will produce an e.m.f. of greater frequency. When applied to a superheterodyne receiver these e.m.fs. produce, in the intermediate-frequency amplifier, corresponding e.m.fs. the frequencies of which, whatever the signal frequency, differ by twice the frequency of rotation of the loop. They can then be separated by fixed band-pass filters, and the means of discrimination between the two senses of rotation is independent of the signal frequency. With continuous waves, this rotation of the loop might be purely physical, but with pulses in which the Fourier component frequencies may cover a range of 10 kc./sec. or more, rotational frequency of the loop must be greater than this figure, or it will not be possible to separate the two components. Such a speed of rotation can only be produced by an artifice. Remembering that a rotating vector can be considered as the sum of two simple harmonic vectors in time and phase quadrature, it is seen that one method is to cause the output from two loops at right-angles to vary sinusoidally but in phase quadrature and to add the two outputs. Grid modulation with variable-mu valves can be made suitably linear and of the correct form, i.e. $\cos \omega t \cos pt$ and not $\cos \omega t\,(1+\cos pt)$, by applying the signal and modulating e.m.fs. to the grids in push-pull and connecting the anodes in parallel. After this stage of modulation (or of superheterodyne detection as it may be called if the matter is regarded from another point of view) the resultant frequencies are passed to any ordinary pulse-receiver and the two component frequencies are separated as if they were different signals. A convenient check is available in that the direction of rotation of the equivalent loop can be reversed, so that either sense of rotation can be received at either frequency. For automatic recording, the upper frequency is chosen arbitrarily, and either right- or left-hand polarization is received by reversing the sense of rotation of the loop. No radio-frequency currents flow through this reversing switch, for the operation is performed in the modulating circuit, and the adjustment of the high-frequency circuits is unaffected.

Such a system has been developed in the Halley Stewart laboratory and is described and investigated in detail below.

§ 4. DESCRIPTION OF THE APPARATUS

The schematic diagram of the apparatus is shown in figure 1. On the left-hand side is the modulating oscillator, the frequency of which determines the frequency of rotation of the effective loop. This frequency is about 33 kc./sec., well above the highest modulation frequencies of the pulse, and the particular value is chosen so that convenient commercial values of capacity and resistance (0·01 μF. and 500 Ω.) may be used in the bridge modulating circuit. Balance to earth is essential throughout the apparatus, so the oscillator is in push-pull disposition as shown. Shunt feed to the anodes permits the bridge circuit to be directly connected to the output of the oscillator. The bias provided by the resistance-capacity circuit R_{12}, C_{12}, is adjusted to give the purest wave-form to the output, since harmonics give rise to unwanted frequencies in the output from the modulator. The inductances L_1 are balanced to earth. The main frequency-adjustment is effected by means of the condenser C_9. The output is taken from across two relatively large condensers C_{10}, of 0·02 μF. capacity to ensure good balance to earth. The resistances R_{10} provide earth connexions for the grids of the valves. This method of coupling reduces the harmonic content of the output to a negligible value.

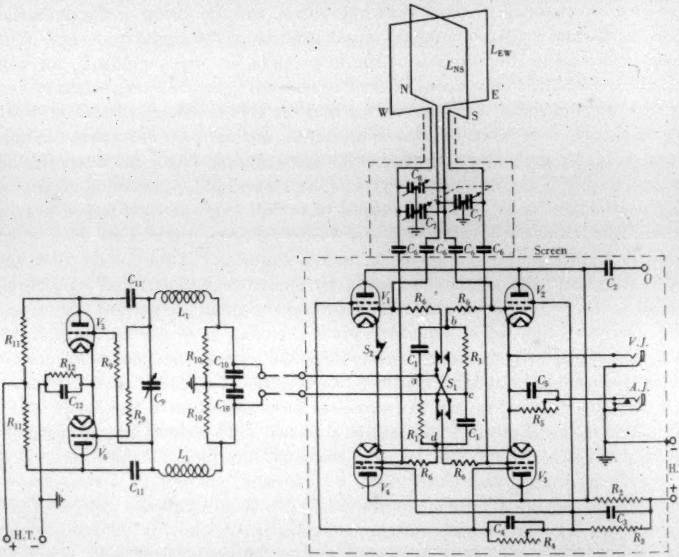


Figure 1. Circular polarization receiver.

V_{ac}, V_{bd}

The output is taken to the two corners a, c, of the resistance-capacity bridge circuit R_1, C_1, R_1, C_1. When the frequency of the oscillator is correctly adjusted, the voltage across the two pairs of diagonals V_{ac} and V_{bd} are equal and 90° out of phase, and the voltages of the opposite points V_a and V_c, or V_b and V_d, of either diagonal are 180° out of phase with reference to earth-potential.

The grids of the valves V_1, V_2, V_3, V_4 are connected each to one corner of the bridge circuit, and the bias on the valves is adjusted by means of the resistance R_5, which is shunted by smoothing-condenser C_5, to obtain a modulating bias of the correct form. The anodes are connected together and the output from the set goes through the condenser C_2 to the first circuit of a superheterodyne receiver normally used for pulse-reception.

Mounted on top of the screened box containing the modulating stage are the mutually perpendicular loops L_{NS} and L_{EW}; four leads are taken from these to the top compartment of the box and connected each to one section of a four-gang condenser having a capacity of 0·0005 μF. per section. Since the moving plates and spindle of this condenser are earthed, the loop circuit is suitably balanced, and trimming-condensers correct this balance. An auxiliary trimming-condenser C_8 ensures that the two loops shall be exactly in tune with each other. Each of the four leads from the loops is also connected to one of the grids of the modulating valves, through a stopping-condenser C_6 of capacity 0·0001 μF. This condenser presents a high impedance to currents of modulation frequency, which would otherwise be short-circuited through the loops. Similarly, resistances R_6 in the leads to the grids from the bridge circuit present a high impedance to radio-frequency currents.

Other necessary components in the set are shown. The jacks *VJ* and *AJ* are for checking the steady bias and the total anode current respectively. The resistance R_2 is the anode feed impedance; C_2 is the output coupling condenser of capacity 0·0001 μF. When the system is operating correctly, currents of modulation frequency should be entirely absent from the anode circuit, so that filtering is not essential. For the same reason, the dropping-resistance for the screen voltages is smoothed by a capacity of only 0·05 μF. A further dropping-resistance R_4 is included in the lead to one pair of valves, so that the maximum amplification of each pair can be adjusted to equality; the current through this resistance is no longer constant and a smoothing-condenser of 1 μF. is required. A switch S_1 reverses the connexions from the bridge circuit to the valves V_1 and V_3 so that, in effect, the direction of rotation of the equivalent loop is reversed. Another switch S_2 opens the cathode circuit of one of the valves and thus unbalances the set sufficiently to produce currents of signal frequency in the anode circuit when desired. The whole apparatus is arranged to be as symmetrical as possible about a vertical axis in order to prevent coupling between the two loops and their amplifiers, and to reduce unavoidable unbalance to a minimum.

The loops are wound with bronze wire, for mechanical reasons. The extra resistance introduced makes the electrical adjustments less critical. The inductances of the two loops are adjusted to equality by means of ebonite bridges which strain together the wires of the loop.

§ 5. THEORY OF OPERATION

Curve *a* of figure 2 shows the variation of the mutual conductance with grid-bias for a typical variable-mu valve (VMS_4). In both pairs of valves in the modulating stage the grids are in push-pull and the anodes in parallel, so to find the curve for the resultant mutual conductance curve *b* must be added to curve *a*. By suitably adjusting the steady bias on the two valves the sum of these two curves, curve *c*, can be made practically linear over a range of ±5·5 V. about the mean bias of −5·5 V. The load impedance of the stage is small so that the mutual conductance

μ
k, E_g
E_1
V, p, t, E, ω

is a measure of the amplification, so that $\mu = kE_g$, where μ is the amplification factor of the stage, k is a constant, and E_g is the voltage on the grid, measured from the point of steady bias. If, in addition, a small voltage E_1 is impressed on the grid, it will produce a voltage $E_1 k E_g$ across the anode circuit. If E_g is replaced by the modulating voltage $V \cos pt$, and E_1 is replaced by the signal voltage $E \cos \omega t$, then the voltage in the anode circuit becomes $kVE \cos pt \cos \omega t$.

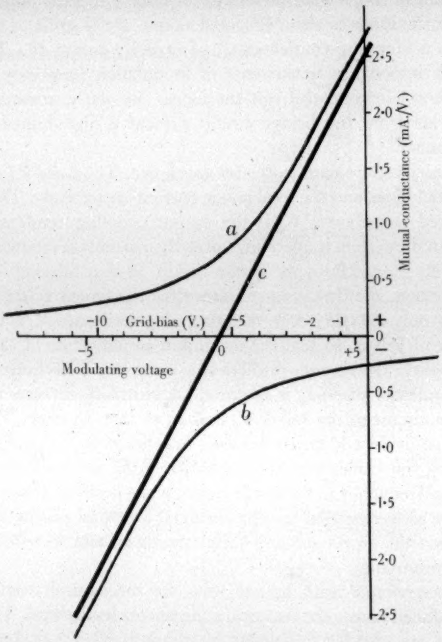


Figure 2.

Now in the other pair of valves the modulating voltage is $E \sin pt$, and if the incident signal wave is circularly polarized, E_1 becomes $\pm E \sin \omega t$, according to the sense of rotation. The resultant voltage across the anode circuit due to all four valves then is

$$kVE (\cos \omega t \cos pt \pm \sin \omega t \sin pt)$$

which

$$= kVE \cos (\omega \pm p)t,$$

i.e. the frequency is either increased or decreased by an amount equal to the modulation frequency, according to the sense of rotation of the incident circularly polarized wave. The signal frequency is not present, and also, if the sense of rotation of the equivalent loop is reversed, that circular component which produced currents at the upper frequency now produces currents at the lower frequency, and *vice-versa.*

In general, whatever the state of polarization of the incident wave or its angle of incidence, the projection of the magnetic vector on to the horizontal plane will be the only portion that can affect the loops, and this projected vector can in turn be resolved into two rotating vectors of opposite sense. One of these rotating vectors will produce in the anode circuit currents of frequency equal to the sum, and the other will produce currents of frequency equal to the difference, of the signal and modulating frequencies. The currents or voltages will be truly proportional to the magnitudes of the two resolved vectors, and a measurement of these currents is a measurement of the state of polarization of the horizontal component of the incident wave.

Since no currents of signal frequency appear in the anode circuit, and the main amplifier is tuned to one or other of the two resultant frequencies, pick-up in the leads from the modulator stage to the main amplifier introduces no errors. It is therefore feasible to locate the loops and modulator at a considerable distance from the amplifier and recording apparatus, if this is desired.

§ 6. ERRORS AND ADJUSTMENTS

Only the right-handed and left-handed rotating vectors of the horizontal component of the magnetic field of the incident wave will be considered, with the usual convention that the sense of rotation is referred to an observer looking along the direction of propagation of the wave before reflection at the ground—i.e. looking downwards. The reference modulation voltage is $\cos pt$ and the reference signal voltage is $\cos \omega t$ so that the right-handed and left-handed rotating vectors are $(\cos \omega t + j \sin \omega t)$ and $(\cos \omega t - j \sin \omega t)$ respectively. At zero time, the signal reference vector makes an angle δ with that loop in which the modulation is $\cos pt$. δ
Conditions, then, are as shown in figure 3*a*.

Owing to errors in the bridge circuit, the two modulating voltages may not be exactly 90° out of phase, so $\sin pt$ is replaced by $\sin(pt+\theta)$. If correctly tuned to resonance, the voltages induced in loop A by the right and left-handed vectors are $\cos(\omega t+\delta)$ and $\cos(\omega t-\delta)$ respectively, while those induced in loop B are $\sin(\omega t+\delta)$ and $-\sin(\omega t-\delta)$ respectively. Even if the two loops are not correctly in tune, but are detuned by equal amounts in the same direction, i.e. if they have the same phase angle, it is still possible to use these two expressions for the induced voltages, if zero time is suitably chosen. But if the phase angles differ by, say, an amount α, then $\sin(\omega t+\delta)$ and $-\sin(\omega t-\delta)$ must be written as $\sin(\omega t+\delta+\alpha)$ and α
$-\sin(\omega t-\delta+\alpha)$ respectively. Another source of error is that either the modulation voltages on the two loops, the voltages induced in the loops for a given signal strength, or the amplifications produced by the two pairs of valves may not be equal. Since the voltage in the anode circuit is the product of all three, it is not possible to separate the various effects. It is sufficient to say that the output from

loop B is $(1-n)$ times that from loop A. Conditions, then, are as shown in figure 3*b*. The resultant effect due to a component of right-hand sense is

$$\cos(\omega t+\delta)\cos pt+(1-n)\sin(\omega t+\delta+\alpha)\sin(pt+\theta)$$

$$\begin{aligned}\text{which}&=\tfrac{1}{2}\left[\cos\{(\omega+p)t+\delta\}+\cos\{(\omega-p)t+\delta\}\right.\\&\qquad\left.+(1-n)\left(\cos\{(\omega-p)t+\delta+\alpha-\theta\}-\cos\{(\omega+p)t+\delta+\alpha+\theta\}\right)\right]\\&=\left(1-\frac{n}{2}\right)\cos\left\{(\omega-p)t+\delta+\frac{\alpha-\theta}{2}\right\}\cos\frac{\alpha-\theta}{2}\\&\quad+\frac{n}{2}\sin\left\{(\omega-p)t+\delta+\frac{\alpha-\theta}{2}\right\}\sin\frac{\alpha-\theta}{2}\\&\quad+\left(1-\frac{n}{2}\right)\sin\left\{(\omega+p)t+\delta+\frac{\alpha+\theta}{2}\right\}\sin\frac{\alpha+\theta}{2}\\&\quad+\frac{n}{2}\cos\left\{(\omega+p)t+\delta+\frac{\alpha+\theta}{2}\right\}\cos\frac{\alpha+\theta}{2}\qquad\ldots\ldots(1).\end{aligned}$$

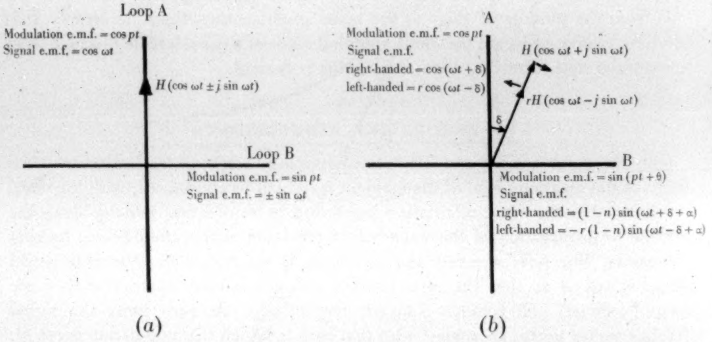

Loop A
Modulation e.m.f. = cos pt
Signal e.m.f. = cos ωt
$H(\cos\omega t \pm j\sin\omega t)$
Loop B
Modulation e.m.f. = sin pt
Signal e.m.f. = ± sin ωt
(a)
A
Modulation e.m.f. = cos pt
Signal e.m.f.
right-handed = cos $(\omega t+\delta)$
left-handed = r cos $(\omega t-\delta)$
$H(\cos\omega t + j\sin\omega t)$
$rH(\cos\omega t - j\sin\omega t)$
δ
B
Modulation e.m.f. = sin $(pt+\theta)$
Signal e.m.f.
right-handed = $(1-n)\sin(\omega t+\delta+\alpha)$
left-handed = $-r(1-n)\sin(\omega t-\delta+\alpha)$
(b)

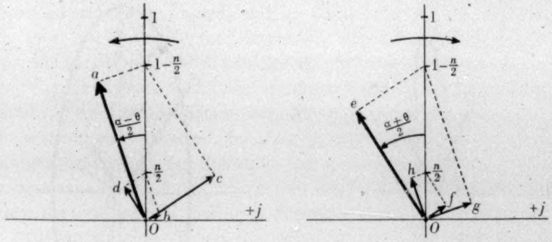
Figure 3.

The resultant effect due to a component of left-hand sense is

$$\begin{aligned}&\cos(\omega t-\delta)\cos pt-\sin(\omega t-\delta+\alpha)\sin(pt+\theta)\\&=\left(1-\frac{n}{2}\right)\cos\left\{(\omega+p)t-\delta+\frac{\alpha+\theta}{2}\right\}\cos\frac{\alpha+\theta}{2}\\&\quad+\frac{n}{2}\sin\left\{(\omega+p)t-\delta+\frac{\alpha+\theta}{2}\right\}\sin\frac{\alpha+\theta}{2}\\&\quad+\left(1-\frac{n}{2}\right)\sin\left\{(\omega-p)t-\delta+\frac{\alpha-\theta}{2}\right\}\sin\frac{\alpha-\theta}{2}\\&\quad+\frac{n}{2}\cos\left\{(\omega-p)t-\delta+\frac{\alpha-\theta}{2}\right\}\cos\frac{\alpha-\theta}{2}\qquad\ldots\ldots(2).\end{aligned}$$

Both right and left-hand components produce effects at the upper and lower frequencies and it remains to be seen just how the various effects can be made obvious. Any apparatus which is sensitive only to a uniformly rotating vector of constant amplitude can have no sense of orientation, in other words the effect produced in the apparatus must not vary as it is rotated about the axis of rotation of the vector—in this case the vertical axis. It is to be expected that errors will affect this axial symmetry. By varying δ in the above expressions the errors should be made apparent.

The vectors representing the terms in the above expressions are shown in figure 4 and their direction of rotation as δ is increased positively is counterclockwise or clockwise according as the terms contain $+\delta$ or $-\delta$, as is indicated in the figure. The experimental procedure is to observe the effects at a single frequency, so the vectors are regrouped as in figure 5 where, in both diagrams, they are of the same frequency and can be added graphically. They rotate in opposite directions as δ is varied, so that this addition has to be performed for each value of δ. Terms containing $\frac{n}{2}\sin\frac{1}{2}(\alpha\pm\theta)$ are neglected as being of the second order of small quantities. To make conditions quite general, an arbitrary state of polarization is assumed, such that the left-hand component is r times the right-hand one. r



$$Oa=\left(1-\frac{n}{2}\right)\cos\left\{(\omega-p)\,t+\delta+\frac{\alpha-\theta}{2}\right\}\cos\frac{\alpha-\theta}{2};\qquad Oe=\left(1-\frac{n}{2}\right)\cos\left\{(\omega+p)\,t-\delta+\frac{\alpha+\theta}{2}\right\}\cos\frac{\alpha+\theta}{2};$$

$$Ob=\frac{n}{2}\sin\left\{(\omega-p)\,t+\delta+\frac{\alpha-\theta}{2}\right\}\sin\frac{\alpha-\theta}{2};\qquad Of=\frac{n}{2}\sin\left\{(\omega+p)\,t-\delta+\frac{\alpha+\theta}{2}\right\}\sin\frac{\alpha+\theta}{2};$$

$$Oc=\left(1-\frac{n}{2}\right)\sin\left\{(\omega+p)\,t+\delta+\frac{\alpha+\theta}{2}\right\}\sin\frac{\alpha+\theta}{2};\qquad Og=\left(1-\frac{n}{2}\right)\sin\left\{(\omega-p)\,t-\delta+\frac{\alpha-\theta}{2}\right\}\sin\frac{\alpha-\theta}{2};$$

$$Od=\frac{n}{2}\cos\left\{(\omega+p)\,t+\delta+\frac{\alpha+\theta}{2}\right\}\cos\frac{\alpha+\theta}{2};\qquad Oh=\frac{n}{2}\cos\left\{(\omega-p)\,t-\delta+\frac{\alpha-\theta}{2}\right\}\cos\frac{\alpha-\theta}{2}.$$

Figure 4.

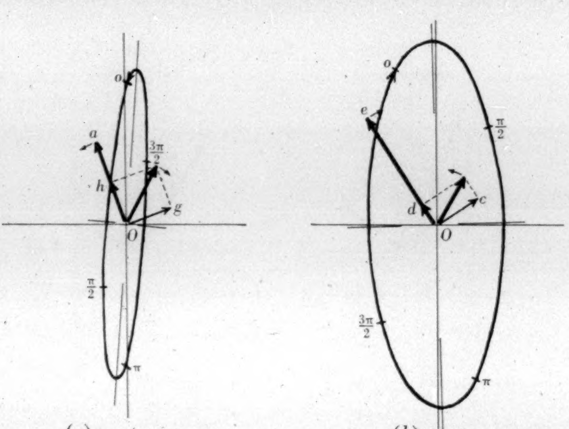
The equations corresponding to figure 5 can be written thus. For figure 5*a* the resultant effect at frequency $(\omega-p)$

$$\begin{aligned}=&\left(1-\frac{n}{2}\right)\cos\left\{(\omega-p)\,t+\delta+\frac{\alpha-\theta}{2}\right\}\cos\frac{\alpha-\theta}{2}\\ &+\left(1-\frac{n}{2}\right)r\sin\left\{(\omega-p)\,t-\delta+\frac{\alpha-\theta}{2}\right\}\sin\frac{\alpha-\theta}{2}\\ &+\frac{nr}{2}\cos\left\{(\omega-p)\,t-\delta+\frac{\alpha-\theta}{2}\right\}\cos\frac{\alpha-\theta}{2}\qquad\qquad \text{......(3).}\end{aligned}$$

For figure 5*b*, the resultant effect at frequency $(\omega+p)$

$$\begin{aligned}=r\left(1-\frac{n}{2}\right)\cos\left\{(\omega+p)\,t-\delta+\frac{\alpha+\theta}{2}\right\}\cos\frac{\alpha+\theta}{2}\\ +\left(1-\frac{n}{2}\right)\sin\left\{(\omega+p)\,t+\delta+\frac{\alpha+\theta}{2}\right\}\sin\frac{\alpha+\theta}{2}\\ +\frac{n}{2}\cos\left\{(\omega+p)\,t+\delta+\frac{\alpha+\theta}{2}\right\}\cos\frac{\alpha+\theta}{2}\qquad\ldots\ldots(4).\end{aligned}$$

The ellipses of figure 5 show the variation of output as the loop assembly is rotated, and noting that δ is measured in opposite directions round the two curves (upon which values of δ are marked) we see that the major axes are almost at right-angles when α is large. As α is decreased, the eccentricity approaches zero and the major axes approach each other, one ellipse having passed through the circular form in the process. Variation of θ and n in the correct direction brings the eccentricity still closer to zero until finally both curves become circles and there is no variation of output as the loops are rotated.



(*a*) $(\omega-p)$ (*b*) $(\omega+p)$

Figure 5.

This, then, constitutes the technique of adjustment and it is evident that no special state of polarization of the testing wave is required. It is inconvenient to retune to find the effects at the two frequencies, but, as has been noted above, the same effect can be produced by altering the direction of rotation of the equivalent loop, i.e. by operating the switch S_1. The set can be adjusted by means of a local oscillator, which does not have to be accurately tuned to the desired signal; nor does it have to be situated at any great distance from the apparatus, several yards being quite sufficient.

The method of operation is as follows. The modulator and main amplifier are tuned to the signal frequency by unbalancing the former, by operating switch S_2, with the modulating oscillator switched off. On closure of S_2 and switching on of the oscillator, the signal can be found at two settings of the heterodyne oscillator, corresponding to the two frequencies $(\omega \pm p)$.* One of these frequencies is selected arbitrarily. The loop assembly is rotated and the positions at which the maxima and minima occur are noted, first with the switch S_1 in one position so that the relay is unoperated, and then in the other so that the relay is operated. If the minima are nearly at right angles to each other in the two cases, the trimming-condenser C_8 is adjusted in such a way as to bring the positions of minima together, and one ellipse should pass through the circular form in the process. Then, by varying condenser C_9 of the oscillator, the variation of output which occurs as the loops are rotated can be decreased, and by adjustment of the resistance R_4 the residual variation can be still further decreased. After a repetition of this series of adjustments, the variation for both settings should be negligible.

This adjustment is made at the high-frequency end of the range of the set and then repeated for the low frequency end; but in this case the loops are tuned to each other, α being thus reduced, by altering their inductance by means of the ebonite bridges, instead of by alterating the trimming-condenser. The residual variation that occurs as the loops are rotated can be reduced below ± 2 per cent, and it is then not symmetrical, owing to detuning of the loop circuits as the leads therefrom twist on rotation. When the adjustment is made at both ends of the frequency range, it is found that, for any setting of the tuning condenser, the error is less than ± 5 per cent; this figure can always be reduced to ± 2 per cent by slight readjustment at the required frequency. Thus it becomes feasible to use the set to determine the state of polarization of the received wave at a series of frequencies in rapid succession, for the whole set can be tuned in the same manner as an ordinary receiver, and the time taken for a complete $\{P', f\}$ curve is not more than twice that required for its delineation by the usual manual methods. If accuracy better than ± 5 per cent is required considerably more time would have to be taken.

There are a number of other errors which may occur, but they can be tested for and eliminated independently. Among these is the effect of harmonics in the modulation stage, with which can be grouped the distortion in the valves. Harmonics can only produce effects at frequencies $(\omega \pm 2p)$, $(\omega \pm 3p)$ ..., and such effects will be filtered out in the intermediate stages of the receiver. They introduce no errors into the measurements, but merely impair the efficiency of the apparatus and raise the general noise-level. Antenna effect may be present, but any voltages due to this cause will presumably be in the same phase on both grids of one pair of valves so that, to an approximation, the resultant modulation will be of the form $\cos \omega t\,(1 + \cos pt)$ and the terms in $(\omega \pm p)$ will have only half the amplitude of those due to a loop voltage of the same magnitude; thus there is convenient discrimination. By careful attention to symmetry in construction and by the use of push-pull modulation, antenna effect can be made negligible.

---

* There are, of course, two other settings associated with the second channel of the superheterodyne receiver.

Mutual coupling between the two loops may exist, but it follows from the treatment of Ratcliffe and White(10) that its effect is to increase the output due to one sense of rotation and to decrease that due to the other by equal amounts, while the output from either sense remains constant as the loops are rotated; the main adjustments are not affected. Denoting the outputs due to the circular components in the absence of this error by R and L and the diminution due to coupling by c, we find that the resultant outputs are $(R+c)$ and $(L-c)$. The sum is $(R+L)$, which is the output from a single loop in the position of maximum signal and is evidently independent of the error. The difference is $(R-L+2c)$ and the output from a single loop in the minimum position is $(R-L)$. By comparing these quantities, the presence of mutual coupling between the loops can be detected, and the test is independent of the state of polarization of the testing wave. The check should be made at the high-frequency end of the range since the mutual effect increases with frequency if, as is most common, the coupling is inductive.

R, L, c

Appendixes A and B deal respectively with the magnitude of the errors to be expected from difference in phase angle in the two loops, and with variation of frequency of the modulating oscillator. The latter effect is not serious but the former is extremely critical and, unfortunately, is present in all types of polarization analysers. The effect may be minimized by damping the loop circuits, but set noise soon limits this artifice. With dipole antennae in place of loops this limitation would not be so serious, on account of the greater signal voltage available, and it might even be feasible to dispense entirely with tuned circuits in the modulator stage. The scheme could then be adapted for the automatic registration of $\{P', f\}$ curves for one component only.

The principal disadvantage of the scheme presented is that it is impossible to obtain any high degree of preselection in the single tuned circuit of the modulator stage, since the resultant frequencies are only separated by twice the modulation frequency, i.e. by about 60 kc./sec. As a result the signal-to-noise ratio is lessened. In order to improve conditions, the modulation frequency might be raised and this would probably be justified, in spite of the slightly greater difficulty incurred in the adjustment of the bridge circuit and in the elimination of the spurious effect due to switching. Care must be taken that neither fundamental nor harmonic frequencies fall within the pass band of the intermediate-frequency filter.

§ 7. ADVANTAGES AND DISADVANTAGES OF THE SYSTEM

Advantages. (i) The system can be adjusted by means of a wave of unknown polarization. (ii) The frequency of the testing-wave need not be accurately equal to that of the required signal. (iii) The means of discrimination between the two senses of rotation is independent of the signal frequency. (iv) The operation of changing the sense of rotation to be received is performed in a circuit carrying currents of fixed and relatively low frequency. (v) The tuning-controls can be so ganged as to

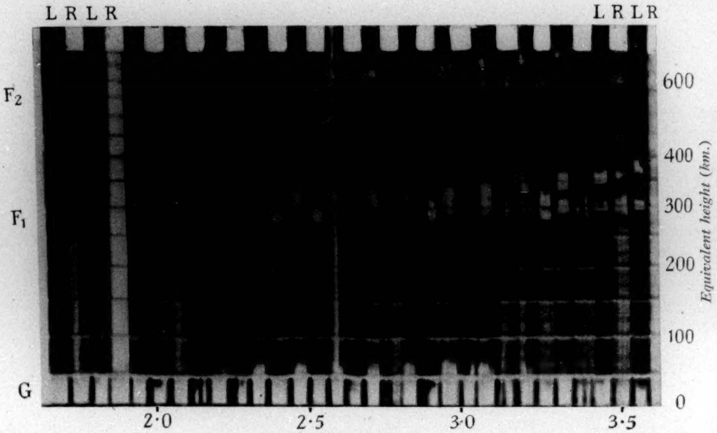


Figure 6.

To face page 1111

permit the reception of a range of frequencies without readjustment. (vi) The system can be used aperiodically. (vii) Physical variation of the circuit constants is, in general, self-compensating.

Disadvantages. (i) The signal-to-noise ratio is lower than in other systems. (ii) The system is somewhat complicated.

§ 8. RESULTS

Figure 6 shows a curve of equivalent height against frequency taken with the apparatus just described. The transmitted frequency has been varied in small steps and alternate frequencies were received with the loop sensitive to one sense of polarization only; the sense is indicated at each step by markings on the upper edge of the diagram. The base line only of the time base is recorded, so that only if the wave is truly circularly polarized does an echo appear for a single sense. Polarization is seldom accurately circular and in general there will be some effect for both senses, as in the case shown in figure 6, but a comparison of the relative amplitudes gives immediately the eccentricity and the sense of rotation.

Owing to facilities being limited, the set has been operated actually inside the laboratory and the distortion of the polarization due to conductors in the building is quite marked at certain frequencies. Thus it is imperative that the apparatus should be capable of being adjusted by means of a wave of unknown polarization. For the same reason we cannot rely on the measured values of eccentricity, but the distortion is never so serious as to lead us to confuse the senses of rotation except for almost linear polarizations.

The record shown was taken in the same time as the usual $\{P', f\}$ curves, i.e. 30 sec. at each frequency, and the set was handled exactly as if it were a normal receiver. No adjustment of the set, other than the necessary tuning was attempted, during the run. The frequency-range covered in the record is from 1·7 to 3·6 Mc./sec. and is roughly the critical region for the Appleton region at this particular time (midnight). It is seen that the two components are very nearly circularly polarized with opposite senses of rotation, at the lower frequencies. The eccentricity departs somewhat from zero as the frequency increases but distortion due to the building becomes apparent at about 3·5 Mc./sec.; even at the end of the record, however, there is a considerable difference between the amplitudes of one component in the two senses.

From the records taken, it seems quite impossible to state accurately the relative intensities of the two components, but the evidence confirms the theoretical result that the component having the lower critical frequency is approximately circularly polarized with a left-hand sense of rotation, and that having the greater critical frequency is also nearly circularly polarized, but with a right-hand sense of rotation.* The records have not been sufficiently numerous to lead to any definite conclusions as to the relative intensities, but with the evidence available from a very extensive series of $\{P', f\}$ curves, it appears that the intensities are quite unpredictable unless the complete $\{P', f\}$ curve for a particular time is available also. The fact that either component may have the greater time delay has long since been accepted. The more complex structure of the ionosphere, which has been revealed as the technique of $\{P', f\}$ measurements has improved, makes it possible that the absorption of either component may increase or decrease with frequency.

---

* Only the range of frequencies above 1·7 Mc./sec. has been under investigation.

Take, for example, the curves of figure 7, which are of a type commonly encountered. The upper curves show the variation of equivalent height of the ordinary and the extraordinary rays for the range of frequencies from 2·4 to 6·0 Mc./sec. Polarization measurements have been made showing that these rays are approximately circularly polarized with opposite senses of rotation, the ordinary being left-handed and the extraordinary right-handed. Below these curves are the corresponding curves for the estimated intensities of the two components. It is seen that the relative amplitudes depend very markedly on the particular frequency under investigation, and either component may be the greater.

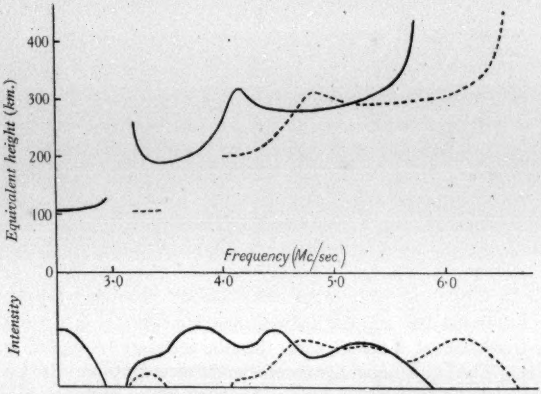


Figure 7. Ordinary ray ——, extraordinary ray - - - - -.

Appleton(14) has observed that the $\{P', f\}$ curves for the two components can be made almost to coincide by moving one of them through about 0·7 Mc./sec. F. W. G. White(15) has given curves showing that the variation of the intensity with time for a fixed frequency follows the same sequence for both components and that the extraordinary ray is less intense than the ordinary ray when referred to corresponding points on the curve of equivalent height. It is seen that this holds also for the $\{P', f\}$ curves, for in figure 8 the curves of figure 7 have been redrawn with the curve for the extraordinary ray displaced by −0·7 Mc./sec., which makes possible a comparison of corresponding points on the two curves. In general, the extraordinary ray is less intense than the ordinary ray at corresponding points, though the difference is less marked at the higher frequencies, because the wave-frequency is far removed from the collisional frequency of the electrons in the layer.

The results indicate, then, that there is ample evidence in favour of the magneto-ionic theory of propagation in the ionosphere, but that deductions from observations made at a single frequency may be misleading unless referred to the $\{P', f\}$ curve appropriate to the time of observation.

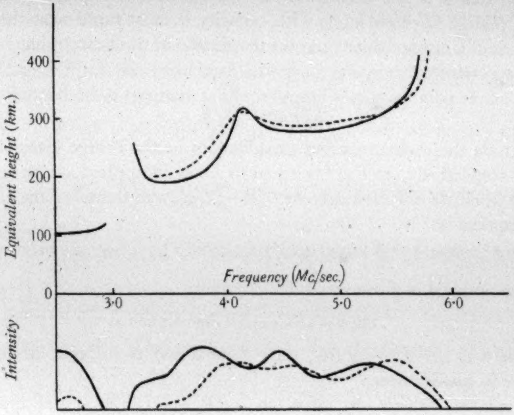


Figure 8. Ordinary ray ——, extraordinary ray - - - - -.

§ 9. ACKNOWLEDGMENTS

The work described in this paper was carried out as part of the programme of the Radio Research Board of the Department of Scientific and Industrial Research. I wish to express my indebtedness to Professor E. V. Appleton for his continued help and encouragement.

APPENDIX A

In a circuit consisting of a resistance R, a condenser C, and an inductance L in series, the current i which flows when an e.m.f. e of frequency ω is applied is given by

$$i\,\{R+j\,(L\omega-1/C\omega)\}=e,$$

and the phase angle α is given by

$$\tan\alpha=(L\omega-1/C\omega)/R.$$

Now if α is small $\tan\alpha=\alpha$

so
$$\delta\alpha=\frac{1}{RC\omega}\frac{\delta C}{C}\quad\text{or}\quad \delta C/C=\delta\alpha/Q,$$

where Q is the voltage step-up at resonance.

It is desirable to limit the error due to the difference in phase angles of the two loop circuits to ±1 per cent and, since the error is $\pm\sin\alpha$, from equations (1) and (2), this requires that α (or in the present notation $\delta\alpha$) shall be less than 0·01 radians. The value of Q will be about 10, whence $\delta C/C = 0{\cdot}001$, i.e. the capacity must be accurate to 0·1 per cent, or the condenser-setting must be accurate to 0·1 of a division. A setting of this accuracy is barely possible except with refined apparatus, so that it is not feasible to rely on a calibration of the condenser dial. The rate of change of phase angle with capacity is most rapid near the resonance point, but it is still appreciable when the amplitude of the current has fallen to one half that at resonance.

APPENDIX B

If we denote the resistances and capacitances in the bridge circuit of figure 1 by R and C respectively, an e.m.f. e equal to $E\sin\omega t$ applied across one diagonal will produce in either branch a current $e/(R-j/C\omega)$, and therefore the e.m.f. across the other diagonal is

$$e\,(R+j/C\omega)/(R-j/C\omega), \quad \text{which} = (R^2-1/C^2\omega^2+j.2R/C\omega)/(R^2+1/C^2\omega^2)$$

and the phase angle θ is given by

$$\tan\theta = 2RC\omega/(R^2C^2\omega^2-1).$$

If $R^2C^2\omega^2=1$, i.e. if $\omega=1/RC$, then $\theta=\pi/2$ and the e.m.fs. across the two diagonals are in quadrature.

If θ is replaced by $(\pi/2-\theta)$, then $\tan\theta=(RC\omega-1)/2RC\omega$ and approximately $\delta\theta=(RC+1/RC\omega^2)\delta\omega/2$. When $\omega=1/RC$ this reduces to $\delta\theta=\delta\omega/\omega$.

As in Appendix A, $\delta\theta$ must not exceed 0·01 radians so that the frequency of the modulating oscillator must be correct and constant to 1 per cent. Under normal conditions this accuracy is easily maintained.

To determine the sense of rotation of the equivalent loop, we note that in figure 1 the e.m.f. e applied to the bridge is that across the diagonal ac and the e.m.f. across the other diagonal is that across bd. Since $\tan\theta$, above, is positive, V_{bd} leads V_{ac} and therefore one side of the equivalent loop passes through W, N, E, S in succession and, if we look down on the loop, this corresponds to a clockwise or right-hand sense of rotation. It will be seen that the operation of the switch S_1 reverses this sense.

APPENDIX C

To investigate the effects of magneto-ionic dispersion on the propagation of a pulse of radio-frequency energy, consider an emitted wave defined by $\cos\omega t f(t)$, where ω is the carrier pulsatance and $f(t)$ is the modulation envelope. If we assume, for simplicity, a symmetrical pulse, $f(t)$ can be expanded in a Fourier cosine series, and if, further, the recurrence pulsatance is taken as unity, $f(t)$ can be written

$$f(t)=a+\sum_{n=1}^{n=a} b_n\cos nt,$$

and the pulse of radio-frequency energy becomes (if we omit the limits of summation)

$$\cos \omega t\, (a + \Sigma b_n \cos nt)$$

which

$$= a \cos \omega t + \tfrac{1}{2} \Sigma b_n \{\cos (\omega + n)\, t + \cos (\omega - n)\, t\}.$$

Now the right- and left-hand circularly polarized components of this wave, namely

$$a\,(\cos \omega t \pm j \sin \omega t) + \tfrac{1}{2} \Sigma b_n \{[\cos (\omega + n)\, t \pm j \sin (\omega + n)\, t] + [\cos (\omega - n)\, t \pm j \sin (\omega - n)\, t]\},$$

travel independently in the ionosphere. Assuming that there is a difference between the paths P_r and P_l of the right and left-hand components, but that these two paths are independent of frequency, the reflected waves arriving at the receiver are

$$a\,[\cos \omega\, (t-\theta) + j \sin \omega\, (t-\theta)] + \tfrac{1}{2} \Sigma b_n\, [\cos (\omega \pm n)\, (t-\theta) + j \sin (\omega \pm n)\, (t-\theta)]$$

and

$$a\,[\cos \omega\, (t-\phi) - j \sin \omega\, (t-\phi)] + \tfrac{1}{2} \Sigma b_n\, [\cos (\omega \pm n)\, (t-\phi) - j \sin (\omega \pm n)\, (t-\phi)],$$

where $\theta = P_r/c$, $\phi = P_l/c$, and c is the velocity of light.

If these waves are received on a simple loop aerial lying in the plane

$$(\cos \alpha + j \sin \alpha),$$

the e.m.fs. induced are given by

$$a \cos \{\omega\, (t-\theta) + \alpha\} + \tfrac{1}{2} \Sigma b_n\, [\cos \{(\omega + n)\, (t-\theta) + \alpha\} + \cos \{(\omega - n)\, (t-\theta) + \alpha\}]$$

and

$$a \cos \{\omega\, (t-\phi) - \alpha\} + \tfrac{1}{2} \Sigma b_n\, [\cos \{(\omega + n)\, (t-\phi) - \alpha\} + \cos \{(\omega - n)\, (t-\phi) - \alpha\}]$$

or

$$\cos \{\omega\, (t-\theta) + \alpha\}\, [a + \Sigma b_n \cos n\, (t-\theta)] = \cos \{\omega\, (t-\theta) + \alpha\}\, f(t-\theta)$$

and

$$\cos \{\omega\, (t-\phi) - \alpha\}\, [a + \Sigma b_n \cos n\, (t-\phi)] = \cos \{\omega\, (t-\phi) - \alpha\}\, f(t-\phi).$$

Thus the envelopes of modulation are reproduced at the receiver after times θ and ϕ and these two pulses are the right and left-hand circularly polarized components predicted by theory, and they exist individually at the receiver. When combined in the receiver they lose their identities, and because of the vector summation the resultant pulse bears little resemblance to the original pulses, unless they are completely resolved.*

Instead of finding the separate effects of the two circularly polarized components, Handel and Plendl combine them into a plane-polarized wave, represented by

$$2a \cos \omega \left(t - \frac{\theta+\phi}{2}\right) \left[\cos \omega \frac{\theta-\phi}{2} - j \sin \omega \frac{\theta-\phi}{2}\right]$$
$$+ \Sigma b_n \cos (\omega \pm n) \left(t - \frac{\theta+\phi}{2}\right) \left[\cos (\omega \pm n) \left(\frac{\theta-\phi}{2}\right) - j \sin (\omega \pm n) \left(\frac{\theta-\phi}{2}\right)\right]$$

---

* The conditions which have been assumed are very unusual in practice but it can be shown(16) that, in the presence of frequency dispersion, expressions similar to the above are derivable, though the phase of the carrier wave is determined by the optical path, whereas the envelope of modulation is reproduced after the equivalent path of the atmospheric waves; the shape of the modulation envelope is not appreciably affected, except under certain extreme conditions.

from which it is apparent that the plane of polarization varies with the side-band frequency. The e.m.f. induced in the aerial becomes

$$2a \cos \omega \left(t-\frac{\theta+\phi}{2}\right) \cos \left\{\omega \left(\frac{\theta-\phi}{2}\right)+\alpha\right\}$$

$$+\Sigma b_n \cos (\omega \pm n) \left(t-\frac{\theta+\phi}{2}\right) \cos \left\{(\omega \pm n) \left(\frac{\theta-\phi}{2}\right)+\alpha\right\}.$$

The summation of the equivalent of this expression is carried out graphically by the authors mentioned, and the possibility of resolving the resultant complex pulse into two overlapping pulses is obscured since the two amplitudes at any point have to be added vectorially. It should also be noted that the polarization varies throughout the overlapping portion, though the beginning and end are circularly polarized with opposite senses of rotation.

The two methods of attack must be strictly equivalent, however, and the above expression readily reduces to

$$\cos \{\omega (t-\theta)+\alpha\} f(t-\theta)+\cos \{\omega (t-\phi)-\alpha\} f(t-\phi),$$

as was found previously.

In determining the time delay of the components of a complex echo, it is useless to take the points of zero amplitude. The echo must first be resolved into two pulses similar in shape to the emitted pulse, and the beginning of these two pulses will give the correct delay for the two components. If measurements can be made from the end as well as the beginning of the pulses, this procedure is not necessary.

REFERENCES

(1) BELLINI. *Electrician*, **86**, 220 (Feb. 18, 1921).
(2) ECKERSLEY, T. L. *Radio Rev.* **2**, 60 and 231 (1921).
(3) APPLETON, E. V. *Proc. phys. Soc.* **37**, 16D (Feb. 1925).
(4) SMITH-ROSE, R. L. and BARFIELD, R. H. *Proc. roy. Soc.* **107**, 587 (1925), and **110**, 580 (1926).
(5) APPLETON, E. V. and RATCLIFFE, J. A. *Proc. roy. Soc.* **117**, 576 (1928).
(6) GREEN, A. L. *Rad. Res. Board of Aust., Report* No. 2 (1932).
APPLETON, E. V. *Nature*, Lond. (Dec. 19, 1931).
(7) HOLLINGWORTH, J. *Proc. roy. Soc.* **119**, 444 (1928).
(8) APPLETON, E. V. and BUILDER, G. *Proc. phys. Soc.* **44**, 78 (1932).
(9) *Wireless World* (July 8, 1932).
(10) RATCLIFFE, J. A. and WHITE, E. L. C. *Phil. Mag.* **16**, 129 (1933).
(11) ECKERSLEY, T. L. *Nature*, Lond. (April 8, 1932).
(12) Discussion on the Ionosphere. *Proc. roy. Soc.* **141**, 697 (1933).
(13) HANDEL and PLENDL. *Elekt. Nachr.-Tech.* **10**, 75 (1933).
(14) APPLETON, E. V. *Proc. phys. Soc.* **45**, 679 (1933).
(15) WHITE, F. W. G. *Proc. phys. Soc.* **46**, p. 805 (1934).
(16) PULLEY, O. O. London Ph.D. Thesis (June 1934).

536.413.2:546.47

MEASUREMENTS OF THE THERMAL EXPANSION OF CAST AND ROLLED ZINC

By H. GOULBOURNE JONES, M.Sc., A.Inst.P.,
County School, Rhyl, N. Wales

Received April 30, 1935

ABSTRACT. The coefficients of expansion of cast and rolled zinc are measured under conditions of steady and of continuously increasing temperature, and marked differences between the coefficients are found to occur. On rolling, the random orientation of crystals in the cast specimen is broken up and most of the long axes are set perpendicular to the plane of rolling. No change is produced in the volume coefficient by rolling. When the zinc crystals are small, the continuous expansion is also small but as the size of the crystals increases, the continuous expansion also increases. Discontinuities in the continuous expansion are traced to the long axis of the crystals.

§ 1. INTRODUCTION

This paper deals with the expansion of small specimens of cast and rolled zinc, as measured with steady and with continuously increasing temperatures. In the latter method, the specimen is heated at a steady rate and simultaneous readings of length and temperature are taken during the heating; observations are also taken during cooling. An optical-lever method measuring down to $8\cdot2\times10^{-6}$ cm. is used to measure the changes of length, while temperatures are measured with a nickel-nichrome thermocouple reading to 0·7° C. The specimens are 0·16 cm. high and can be heated to 405° C. with safety.

The scope of the investigation can best be seen from the following summary. (i) The expansion of the zinc is compared with that of glass, whose expansion is compared with that of silica. The glass has a constant coefficient of $8\cdot6\times10^{-6}$ up to 240° C. During cooling a slightly lower coefficient is obtained. (ii) With steady temperatures, well annealed cast zinc gives a coefficient of $30\cdot2\times10^{-6}$ up to 250° C.; this shows that a random orientation of crystals exists. (iii) Measurements on cast zinc at continuously increasing temperatures reveal sharp discontinuities at about 200° C. in the slope of the {expansion, temperature} graph. Heating the specimen to 400° C. before the experiment moves the discontinuity to a lower temperature around 130° C. After quenching of cast zinc, a slight negative movement is obtained at lower temperatures followed by a slight expansion. On annealing, the expansion increases with the amount of annealing. Thus in the small specimens used, the expansion with continuously increasing temperatures is influenced by the state and growth of the zinc crystals. (iv) The continuous-cooling curves give reliable information and from them the coefficients of expansion of annealed and unannealed zinc can be compared. (v) Steady-temperature measurements with rolled zinc 0·03 cm. thick show that 90 per cent of the short axes of zinc crystals are in the direction of rolling and 80 per cent of the short axes are in the plane but perpendicular to the direction of rolling. (vi) Continuously increasing temperatures with the rolled zinc show that the discontinuities in the expansion are associated with the long axes of the crystals. (vii) The volume coefficient of the zinc is found to be unaltered by rolling.

§ 2. APPARATUS

The inverted form of furnace and optical system used have already been described[1] and it will be sufficient to describe the mounting of the specimen, figure 1. A tripod *D* rests with its outer legs on glass and its centre leg on the specimen. A microscope cover 0·01 in. thick was broken into three pieces and a piece was placed under each foot of the tripod to prevent the foot from sinking into the zinc at higher temperatures. Light passes up the silica tube *T* through the glass window *A* and is reflected by the mirror *C* back approximately along its own path.

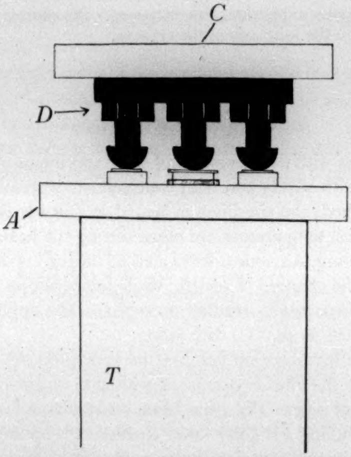


Figure 1. The mounting of the specimen.

The legs of the tripod are made as nearly equal as they could be made by adjustment with vernier calipers, and then an expansion experiment is performed without any specimens. A very slight adjustment of one of the legs is sufficient to eliminate the expansion of the tripod.

Figure 1 shows a rolled specimen 0·030 cm. thick under the centre leg of the tripod. The shaded strip under the centre leg is a small piece of commercial zinc in which are cut two slits; two pieces of rolled zinc are pushed into the slits and so held on edge. The microscope cover piece is now placed across the vertical zincs and the foot of the tripod rests on the glass cover. Strangely enough this gives an exceptionally steady system that is not influenced by outside disturbances.

In calculating the coefficient $(I/L)\,dL/dT$ for continuously increasing temperatures, a minimum overall accuracy of 1 in 50 has been maintained, but for steady temperatures a much greater accuracy is reached, the expansion being measured with an error never greater than 1 in 150, while corresponding temperatures are measured to 1 in 160. The coefficient obtained is in turn averaged out on the {coefficient, temperature} curve.

§ 3. RESULTS FOR GLASS

Focusing difficulties prevented zinc from being compared directly with silica, so the expansion of the zinc was compared with that of glass which in turn was compared with that of fused transparent silica. The coefficient of expansion of silica is quoted as $0{\cdot}54\times10^{-6}$ up to 1000° C.

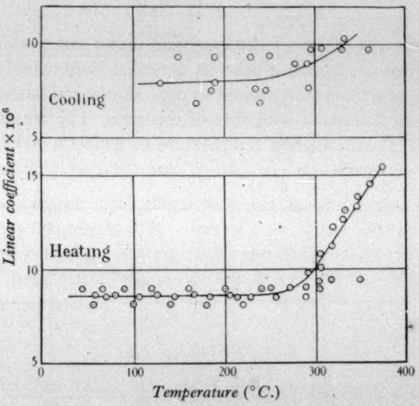


Figure 2. The coefficient of expansion of glass.

After the glass in the furnace had been heated a number of times, observations were taken, and figure 2 gives the coefficient obtained on four occasions with continuously increasing readings of expansion and temperature. This coefficient is actually $(I/L)\,dL/dT$, i.e. the rate of change of length with temperature at a particular temperature. Excellent agreement is obtained for the heating curve, but the cooling curve shows fatigue, as might be expected. At a later date the glass was again compared with silica, the temperature and expansion being kept constant for at least five minutes before a reading was taken. The coefficient thus obtained for steady temperature is identical with the heating coefficient shown in figure 2 for continuously increasing temperatures. Table 1 gives the heating coefficient $(I/L)\,dL/dT$ for glass.

Table 1

Temperature (° C.)	$\frac{I}{L}\frac{dL}{dT}\times 10^6$
30	8·55
200	8·63
300	10·10
380	16

Various rates of heating, such that from 38 to 64 minutes were taken in heating through 310° C., have been used in the continuously increasing temperature measurements, but the coefficient is independent of the rate of heating. The glass was tested at frequent intervals and it is quite certain that the anomalous effects described later are not due to the glass.

§ 4. RESULTS FOR CAST ZINC

To measure the expansion of the cast zinc under conditions of steady temperature, a specimen 0·1780 cm. thick was annealed inside the furnace at 350° C. for 56 hours. The specimen was allowed to cool slowly inside the furnace, and its expansion was then compared with that of the glass. The length of the zinc increased up to 403° C., the highest temperature to which it was heated. Figure 3 gives the corrected coefficient $(I/L)\,dL/dT$ for the cast zinc and on the same figure is shown the coefficient worked from the results of Owen and Yates$^{(2)}$. In using the results of Owen and Yates, quoted for the axis of the zinc crystals, the average linear coefficient for the zinc is taken as $\frac{1}{3}$ (coefficient long axis + 2 × coefficient short axis). Exceptionally good agreement is obtained in the results including the upward trend above 260° C. The agreement indicates the reliability of the method and also indicates a random orientation of zinc crystals inside a well annealed specimen of cast zinc.

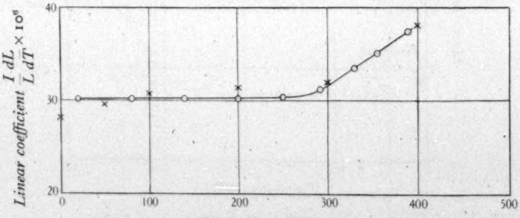


Figure 3. Annealed zinc; × Owen and Yates; ⊙ present work.

Table 2 gives the mean coefficient over the range extending from 20° C. to

100° C. as derived from the present measurements and from the results of Gruneisen and Goens(3).

Table 2

Mean coefficient for cast zinc from 20° C. to 100° C. × 10^6		
Present measurements	Owen and Yates	Gruneisen and Goens
30·3 ± 0·2	29·8	30·7

§ 5. CONTINUOUSLY INCREASING TEMPERATURES

Before the expansion of the cast zinc was examined under conditions of steady temperature, a specimen was examined under continuously increasing temperature. Sharp changes occur in the slope of the {expansion, temperature} curves; so sharp indeed are these changes that they constitute well marked discontinuities. Table 3 indicates their main aspect. From the table it will be seen that heating the specimen immediately before the test has the effect of moving the discontinuity to a lower temperature, while continued heating causes the discontinuity to disappear. It has been found that these discontinuities do not depend on the rate of heating or on changes in the rate of heating.

Table 3

Curve no.	Date	Whether previously heated	Observed heating	Dis-continuity at ° C.	Coefficient up to discontinuity × 10^6
V	May 16	—	3	210	29·1
VI	,, 19	—	4	192	29·6
VII	,, 22	Yes	6	105	24·3
VIII	,, 23	Yes	8	155	25·7
IX	June 13	—	15	—	22·6
X	,, 26	—	16	—	24·3

§ 6. HYPOTHESIS

It is suggested that using very small specimens gives rise to an anomalous effect that would not be detected in large specimens, and that the growth of zinc crystals during a heating experiment may affect the apparent expansion of the zinc. The following experiments have been carried out to test this hypothesis. A specimen was kept inside the furnace at 350° C. for 20 minutes; the top of the furnace was quickly removed and the specimen was quenched in tap water. On being quenched the crystals are subjected to severe strain and will be broken up into smaller crystals. The day after quenching, the expansion of the specimen was compared with that of glass by the continuously-increasing-temperature method. The result is shown in figure 4 as curve XXXI. The movement in the eyepiece is exceptionally small and in addition gives a small negative expansion at lower temperatures. This negative expansion is similar to that obtained at lower temperatures for the first direction (see later) of the rolled zinc. Hence the first direction of the rolled zinc—the direction perpendicular to the plane of rolling—contains a large percentage of small crystals.

The following day, a second experiment was performed on the same specimen without further annealing. The result is shown as curve XXXII, figure 4. These two curves would not be expected to be identical, because the specimen has been annealed to some extent during its first expansion experiment. The general features of the first curve, including the negative movement at lower temperatures, are, however, reproduced in curve XXXII. The linear coefficient, corrected for glass, over most of the heating range is 10×10^{-6}.

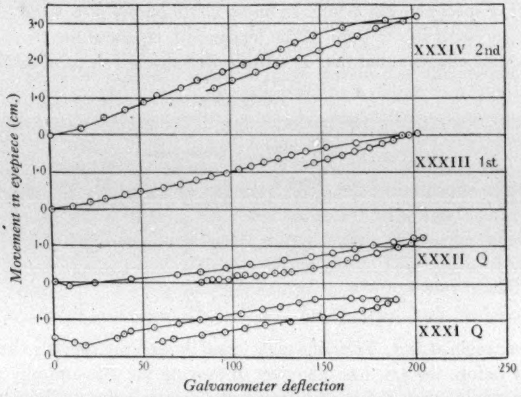


Figure 4. Cast zinc: continuously-increasing-temperature curves obtained after varying heat treatment.

If the specimen is now annealed at a fairly high temperature, some of the crystals will be given a chance to grow at the expense of others and the expansion curve should, on this hypothesis, alter. Accordingly, the specimen was placed inside the furnace without the tripod, mirror, etc., and kept at a temperature of 350° C. for 14 hours. This time does not include the time (about 40 minutes) taken to heat the furnace or the time taken for cooling. During cooling, the 1-inch observation hole in the bottom of the furnace was plugged with asbestos, otherwise strain might be set up in the crystals causing them to be broken once more.

The specimen was subjected to increasing amounts of annealing at 350° C., the expansion being measured between each annealing. Some of the experimental curves are shown in figure 4, while table 4 gives the data. Curve XXXVII is the steady-temperature curve which has already been quoted in the results for cast zinc.

The coefficient $(I/L)\,dL/dT$, corrected for glass, for all the above curves is shown in table 4. This table shows clearly that on quenching the small crystals are in a state that retards expansion. The coefficient of expansion increases with the

Table 4

Curve no.	Date	Type	Total hours annealing	Heating current	Coefficient $\times 10^6$ at 150° C.	Symbol in figure 5
XXXI	Jan. 24	After quenching	—	1·2	17·5	⊙
XXXII	,, 25	After quenching	—	1·2	15	⊡
XXXIII	,, 31	After first annealing	14	0·6 to 1·6	20·4	×
XXXIV	Feb. 7	After second annealing	26	1·2	31·1	⊠
XXXV	,, 15	After third annealing	43	1·0	35·5	⊞
XXXVI	,, 21	After fourth annealing	56	1·4	41·2	⊕
XXXVII	,, 22	Steady	56	—	31·2	

first and second annealings, when the crystals grow, and approaches the coefficient relating to steady temperature. After the third and fourth annealings, the coefficient of expansion exceeds that related to the steady state and it is suggested that these curves correspond to the actual growth of crystals. This view is supported by the way in which the fourth curve falls to the steady-temperature curve at 220° C., as if a saturation point had been reached.

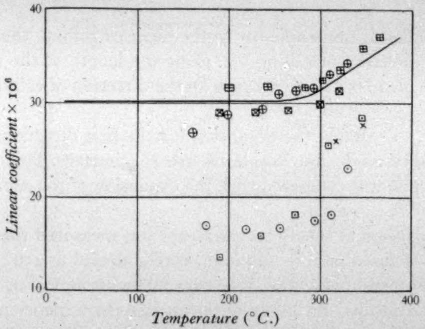


Figure 5. Cast zinc: cooling values during treatment. The continuous line is the steady heating curve.

Twenty days after the steady-temperature curve had been obtained, another continuous-expansion experiment was performed on the same specimen without further annealing. Up to 152° C., the expansion was similar to that after the first annealing; above 152° C., the expansion was similar to that after the fourth annealing, the discontinuity in this curve being identical in nature with the discontinuities already mentioned. On cooling, the coefficient was larger than that obtained in an experiment immediately preceded by annealing.

Figure 5 shows all the cooling-coefficients for the heat-treatment tests of table 4.

The outstanding point of this figure is the exceptionally good agreement obtained between the coefficients for the different amounts of annealing, except in the cases of the quench and first annealing. The full line through these points is the heating-coefficient found with steady temperatures; hence the cooling-coefficient with continuously changing temperature gives reliable information for the cast zinc provided the specimen has been annealed for at least 26 hours.

These results indicate that the continuous expansion of cast zinc containing very small crystals is small also, but as the crystals grow the expansion increases. If the specimen is not annealed before the experiment, the expansion is influenced by the growth of crystals during the experiment.

§ 7. ROLLED ZINC AT STEADY TEMPERATURES

A cast zinc specimen 0·4 cm. thick was rolled down to 0·030 cm., the same end being inserted between the rollers for each rolling. Between each rolling the specimen was heated on the forge, allowed to cool for a few seconds, and then gently rolled. There was sufficient oil on the rollers to quench the small specimen used. An interval of two months was allowed to elapse before expansion experiments were performed on the rolled zinc, so that any periodic fluctuations set up by the working would have time to disappear. The expansion of the rolled zinc was examined in three directions at right angles to each other called the *first, second, and third directions.*

The first direction is perpendicular to the plane of rolling, the second is in the plane and in the direction of rolling, i.e. along the length of the rolled strip; and the third is in the plane but perpendicular to the direction of rolling, i.e. across the rolled strip.

First direction. To obtain the expansion for the first direction, small pieces of the rolled zinc were made into a pile of five specimens 0·1501 cm. thick. The expansion of this pile was compared with the expansion of the same pieces of glass that were used for the cast zinc.

Before the expansion at steady temperatures was measured the specimens were heated a number of times and, in addition, were annealed at 350° C. for 22 hours. The expansion was measured on the three days following annealing. Curve XXXX, figure 6, gives the results. At lower temperatures the expansion was slight but negative, and irregular in that the results obtained on different occasions did not agree, and also irregular in that lagging took place. Around 100° C. the coefficient became positive and from 200° C to 300° C. fair agreement was obtained between five readings.

In figure 7 the coefficient $(I/L)\, dL/dT$ is shown for this work and also for the long and short axes of the zinc crystals from the results of Owen and Yates(2). The agreement indicates that by the rolling of the specimen down to 0·030 cm. a large percentage of the zinc crystals have been turned, so that their long axes are now perpendicular to the plane of rolling. At lower temperatures the expansion is irregular, the irregularity probably being due to the fact that the crystals in this direction are very small (see the results for continuously changing temperature, quoted later).

Second and third directions. The expansion in the second direction measured with steady temperatures is shown in figure 6, and a portion of the cooling curve also is shown. The highest reading on this curve was kept steady for 37 minutes. Up to 240° C. the coefficient is constant at $20{\cdot}5 \times 10^{-6}$, as shown in figure 7. Also in figure 7 will be seen the coefficient $(I/L)\, dL/dT$ worked out for the short axis of the zinc crystals from the results of Owen and Yates⁽²⁾. The coefficient for the second direction lies close to the coefficient for the short axis, and if the two temperatures of 130° C. and 170° C. (which appear from the graph to be reasonable or average temperatures) are taken for computing the percentage of short axes in the

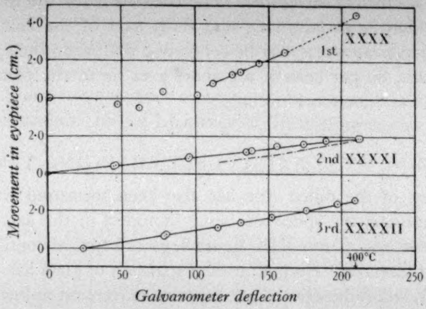


Fig. 6. Rolled zinc compared with glass. Steady temperatures.

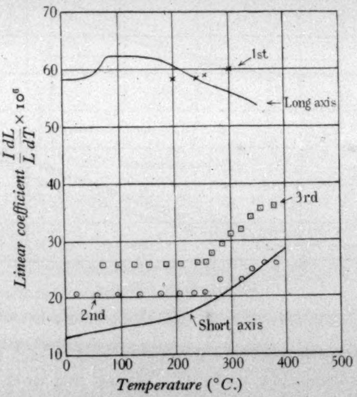


Fig. 7. Coefficients of rolled zinc and long and short axes of zinc crystals.

second direction, then the percentage of short axes is 89 at 130° C. and 91 at 170° C. Thus, when zinc is rolled down to 0·030 cm., 90 per cent of the short axes of the crystals lie in the direction of rolling.

The steady-temperature expansion for the third direction also is shown in figure 6 and the coefficient in figure 7. The coefficient for the third direction lies between the coefficients for the long and short axes of the zinc crystal and, an average being taken as above, it may be concluded that when the specimen is rolled down to 0·030 cm., 80 per cent of the short axes lie in the plane of rolling but perpendicular to the direction of rolling.

§ 8. ROLLED ZINC AT CONTINUOUSLY CHANGING TEMPERATURES

The expansion of the rolled zinc has also been measured with continuously changing temperatures, the specimen being mounted in the same manner as for the steady temperatures. Curve XIX, figure 8, gives a typical continuous-expansion curve for the first direction. Even after the expansion of glass has been allowed for the curve, XIXA, still indicates a slight negative movement at lower temperatures. The minimum on this curve has been obtained on eleven occasions, all with steady heating-current ranging from 0·6 to 1·5 amperes, and in each case the temperature of the minimum has been close to 87° C. During this work the specimens were piled in a different order, allowed to stand for some days, heated to 400° C. and allowed to cool slowly before an expansion experiment, but still the minimum occurred. Now a minimum similar to this occurs in cast zinc immediately after quenching when the zinc crystals are small; hence it is inferred that after rolling also, the crystals are small. Since this minimum is only found in rolled zinc for the first direction, the suggestion is put forward that it is the long axes of the zinc crystals that are associated with the discontinuities obtained for the cast zinc. This

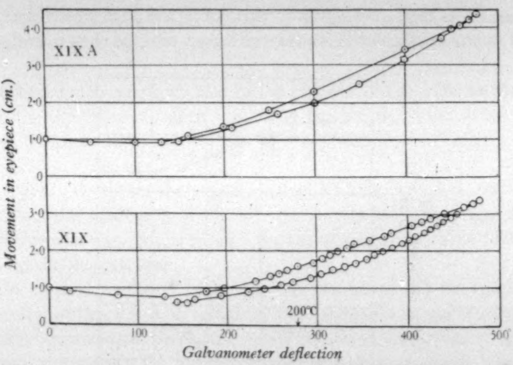


Fig. 8. Typical of 1st direction of rolled zinc. Top curve is corrected for the expansion of glass.

view is supported by the fact that very little difference is found between the heating and cooling coefficients for the second direction containing 90 per cent of short axes.

The third direction of the rolled zinc contains 20 per cent of long axes, and when it is heated bumps are obtained in some three of the continuous-expansion curves. However, the next two heatings of the specimens, without annealing, did not show the bumps but gave constant heating-coefficients of 24·3 and $25{\cdot}1\times10^{-6}$. The cooling-coefficients for all these five curves are in excellent agreement with each other.

In the tests of heat treatment for cast zinc, it was found that the cooling curves gave reliable information. Hence by means of the cooling curves an average coefficient for the rolled zinc can be worked out at any temperature by adding the coefficients for the three directions and dividing by 3. Figure 9 shows the average coefficient for the rolled zinc found in this way.

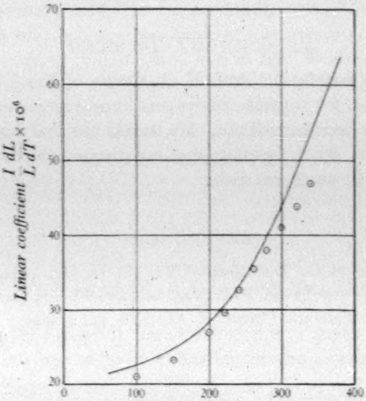


Figure 9. Unannealed zinc: —— cast; ⊙ rolled.

Figure 9 also shows the average coefficient obtained with continuous temperatures for the three directions of the unannealed cast zinc, and this is practically a cooling-curve. The agreement between these two curves for continuous temperatures is too good to be accidental even though they are both very different from the steady-temperature curve of figure 5. It is suggested that the continuously-changing temperature curves are valid for unannealed zinc while the steady-temperature curve applies to annealed zinc. It follows from figure 9 that the volume coefficient of the zinc remains unaltered after the specimen has been rolled down to 0·030 cm.

The expansion of a metal is usually represented by the equation

$$L=L_0\,(1+\alpha T+\beta T^2+\gamma T^3+\delta T^4),$$

where α, β, γ and δ are constants for the metal. On differentiation

$$\frac{1}{L}\frac{dL}{dT}=\alpha+2\beta T+3\gamma T^2+4\delta T^3.$$

This last expression is what has been called the "coefficient" in this paper. Hence the values of the constants can be found.

	$\alpha\times 10^6$	$\beta\times 10^9$	$\gamma\times 10^{12}$	$\delta\times 10^{15}$
Cast zinc unannealed (author)	27·84	−47·52	178	− 18
Cast zinc annealed (author)	29·65	10·84	− 68·4	129
Annealed zinc (Owen and Yates)	28·04	23·39	− 81·11	118·2

§ 9. ACKNOWLEDGMENTS

I am extremely grateful to Prof. E. A. Owen of the University College of North Wales, Bangor, for constant advice and encouragement, without which the work could not have been carried out. My thanks are also due to the Headmaster of the County School, Rhyl, for placing at my disposal the facilities of the school laboratories, where the work was done.

REFERENCES

(1) Jones, H. G. *J. sci. Instrum.* **11**, No. 10, Oct. 1934.
(2) Owen and Yates. *Phil. Mag.* Series 7, **17**, 113 (Jan. 1934).
(3) Gruneisen and Goens. *Z. Phys.* **29**, 141 (1924).

OBITUARY NOTICES

FRANCIS JOHN WITTS

MR FRANCIS JOHN WITTS was born at Newport, Monmouthshire in September 1878. He was apprenticed to Messrs Dunscombe and Co., opticians, of Bristol. He came to London in 1908, and in the same year was elected a member of the Optical Society. In 1914 he opened a business as an ophthalmic optician in the Buckingham Palace Road, and was particularly successful in dealing with patients requiring difficult prescriptions. He was an ingenious man, and invented a rimless pince-nez fitting in which the metal work or clip was made in one piece. He specialized in the drawing of fine wires, and in 1925 invented* a form of double jewel draw-plate with a view to improving the quality of the wires drawn. Mr Witts had a kind and generous disposition. He died on March 4, 1934.

---

* British patent specification No. 257,366.

C. F. B. KEMP, A.R.C.S., B.Sc., D.I.C.

THE death of Charles Frederick Barnard Kemp in a motor accident on July 21, 1934 at the very early age of 31 years has terminated a career which promised to be very successful.

Mr Kemp entered the Royal College of Science in 1921 and took the associateship course in physics. He was awarded the A.R.C.S. (1st Class) in 1924, and took his B.Sc., London, degree (honours physics) in the following year. He was appointed to the Research Department, Woolwich Arsenal in 1925, and transferred to the A.D.E.E. in 1929, where he remained until his untimely death.

During the last few years he had undertaken pioneer work on the analysis of aeroplane noise, and he attacked this most difficult problem with great energy and resource. Some of the experimental results have appeared in the two contributions which he made to the *Proceedings of the Physical Society*, and one has only to glance through these papers to realize what pains he had taken to master the difficult technique of outdoor sound-intensity measurements.

He had been a member of the Physical Society, as Student and Fellow, since 1921.

SIR ARTHUR SCHUSTER, F.R.S.

ARTHUR SCHUSTER was born at Frankfurt-on-Main on September 12, 1851. His family was of Jewish origin and had been settled in Frankfurt for a century or more when, in the middle of the eighteenth century, the firm of Gebrüder Schuster trading in English cotton and other goods was founded. When Napoleon after the conquest of Prussia in 1806 confiscated the goods of the firm as English, several members of the family moved to England, and Manchester became the firm's headquarters, with Frankfurt, London and Bradford as branches. In several of them banking business grew up automatically, and by the middle of the century the Frankfurt business was almost exclusively banking, with Arthur Schuster's father Francis Joseph (born 1823) as head. He had married in 1849 Marie (born 1830), daughter of Hofrat Pfeiffer, one of the founders of the Hofbank at Stuttgart. By this time many members of the family, including all those in England, were Christian, and Francis and Marie became Christian early in their married life and their children were brought up in that faith. She was musical and as a member of the local choral society took part in its sacred concerts. Arthur, the second son, began school at 5 and entered the Gymnasium at 12. Although from his school reports he appears to have made satisfactory progress, he was far from happy in his school life and left at $16\frac{1}{2}$ without regret. In the autumn of 1868 he was sent, as his elder brother Ernest[*] had been the year before and as his younger Felix[†] was the year after, to live with the Protestant Pastor Goetz at Geneva. Here he studied French, and during his two years' stay attended lectures in history, commercial geography, mechanics, physics, chemistry and anatomy at the Academy, and learnt the routine work of the Observatory. He regarded this as one of the happiest periods of his life, and left Geneva with regret in May 1870 for Manchester, where his father had the previous year replaced in the business an uncle who had died. In due course the family was naturalized, and in 1873 Francis Schuster moved to London to join the firm of Schuster, Son and Co., merchants and bankers, Cannon Street, which had been founded in 1845, and part of whose business was taken over by the Union Bank of London in 1887.

In the autumn of 1870 Arthur commenced work in the Manchester office of his father's firm. He had, however, set his heart on a scientific career, and he began to attend the evening lectures on Chemistry given by Prof. Roscoe in the Owens College. At the end of the course, in April 1871, he was awarded the first prize, and his mother who, although then blind, had discovered how unhappy he was in office work, induced his father to consult Prof. Roscoe as to the feasibility of Arthur taking up a scientific career. As a result he entered the day classes at the Owens College in October, to study mathematics under Prof. Barker and physics under Prof. Balfour Stewart, and to commence spectroscopic research under Prof. Roscoe. At the end of the session he was awarded the first prizes in the mathematics, natural philosophy and practical physics classes, and the second prize in the extra physics class. His father was quite satisfied as to his career and took great pride in all his future work. In the summer he communicated some of the results of his research to the British Association at Brighton, and to the Royal Society. The winter he spent at Heidelberg under Professors Kirchhoff, Bunsen and Königsberger and in the summer of 1873 the degree of Ph.D. of that University was conferred on him. In the autumn he was appointed honorary demonstrator in the Physics Department of the Owens College, and commenced a research on the validity of Ohm's law with Prof. Stewart in the new college buildings in Oxford Street. In 1874 at the close of the session he spent a few weeks in Prof. Weber's primitive laboratory at Göttingen, the autumn in Prof. Helmholtz's laboratory at Berlin, and Christmas at his father's house at Hampstead. Early in 1875 he was asked to join and ultimately to take charge of the Royal Society expedition to Siam to observe the solar eclipse of April 6. On the return journey he spent two months in a tour across the Himalayas into Kashmir, and in November he resumed the demonstratorship and his optical work at the Owens College, and lectured to advanced students on Maxwell's "Electricity and Magnetism" which had appeared two years previously.

---

[*] Afterwards K.C. and an authority on international law.

[†] Afterwards Sir Felix Schuster, Bart., of the Union of London and Smiths Bank.

In May 1876 he visited Maxwell at Cambridge with a view to beginning measurements of the magnetic properties of rocks, but after discussing with him the apparatus which would be required he decided to continue his spectroscopic work. He entered the Cavendish Laboratory in October and was admitted a fellow commoner of St John's College in December. In 1878 he went to Colorado to take observations during the solar eclipse of July 29, and in 1879 was elected a Fellow of the Royal Society. When after the death of Maxwell Lord Rayleigh became Professor at the beginning of the Lent term, 1880, Schuster joined with him in a re-determination of the British Association ohm, the value of which was then uncertain to about 4 per cent. In the following Michaelmas term he lectured on radiation to post-graduate students. In 1881 he was elected to the newly constituted chair of applied mathematics at the Owens College, which had become one of the constituent colleges of the Victoria University. He there commenced the researches on the passage of electricity through gases which he described in his Bakerian Lectures to the Royal Society in 1884 and 1890. In 1882 he went with Abney to Egypt to photograph the corona during the solar eclipse of May 17, and in 1886 with L. Darwin and Maunder to the West Indies for that of August 29.

In 1887 he married Cary, the eldest daughter of George Loveday of Wardington, Banbury, Oxon., and in 1888 succeeded Balfour Stewart as Professor of Physics at Manchester.

The great increase in the number of students taking physics necessitated the addition of three rooms to the three which had served previously as the physical laboratories, and in 1898 an independent building for physics, to the design of which he gave much time and thought, was commenced. It was opened in 1900.

When the Charter of the University of Manchester was granted in 1903 Schuster became Dean of the Faculty of Science and representative of the University on the Education Committee of Manchester. His father died in 1906, and he resigned his

Professorship in 1907, being succeeded by Rutherford. Already a good painter in water colours, he now took up oil painting as a recreation and in the course of a tour of India in 1908 he produced a number of oil paintings of Indian scenery.

From 1912 to 1914 he was President of the Physical Society. From 1912 to 1919 he was Secretary and from 1920 to 1924 Foreign Secretary of the Royal Society. In 1913 he removed from Manchester to Twyford, Berks. He was on his way to the Crimea for the solar eclipse of August 21, 1914 when the great war broke out and he only succeeded in returning from Constantinople via Egypt at the end of September. Next year he was President of the British Association at the meeting in Manchester, and on the formation of the International Research Council in 1919 he was elected Secretary.

He was knighted in 1920 and in 1924 served on the Royal Commission on the Universities of Oxford and Cambridge.

In 1923 he lost the sight of one eye through an accident. It was a considerable shock to his general health and well-being, and he had to abandon his recreation of painting. A severe illness in the winter of 1927–8 left him much weakened, and he became a greater invalid as the years progressed. He died at his home at Twyford on October 14, 1934, and a memorial service was held in St James's Church, Piccadilly on October 18.

Lady Schuster, one son and four daughters survive him.

When Schuster became a student under Roscoe he was already searching for harmonic relations amongst the spectral lines of substances, and it was not till ten years afterwards that he showed that such relations were only accidental.

When 25 years later Rydberg had divided the lines of a spectrum into series, known respectively as the P, S and D series, Schuster discovered independently of Rydberg the equality of the wave-number of the first line of the P series and the difference between the limits of the S and D series, a relation which is now included in the modern theory of series spectra.

On the advice of Roscoe he commenced in the Physical Laboratory an experimental investigation into the origin of the second or band spectra of gases under low pressure, and by 1872 had concluded that the band spectrum of hydrogen was due to hydrocarbon impurities derived from the lubrication of the taps of the containing tubes and that of nitrogen was due to the oxides of nitrogen. Widening of the lines of the spectra of gases he ascribed a year later to increase of pressure exclusively, and not to increase of temperature. During his tenure of the physics demonstratorship under Balfour Stewart his experimental tests of Ohm's law led to doubts as to its validity. In 1876 he solved the question of the origin of the motion of the vanes in a radiometer exposed to radiation by surrounding them with a transparent bell, which he showed rotated in the opposite direction to the vanes, thus proving that both motions were due to the action of the residual gas. At Cambridge he continued his investigations of the spectra of metalloids and was led to consider the wider question of the mechanism of the passage of electricity through gases. With Lord Rayleigh he determined the British Association standard ohm in absolute units by the rotating-coil method, and found it 1·1 per cent too small. For several years after his return to Manchester his principal experimental work was on the electrical conductivity of gases, which he held to be ionic in character. He showed that a solid obstacle in the cathode dark space of a vacuum tube cast shadows on both cathode and anode, and on deflecting the rays from the cathode by means of a magnetic field he concluded that these rays consisted of particles of nitrogen. The potential at a distance x from the cathode he expressed as $V(1-e^{-kx})$, where k was a constant. His Bakerian lectures of 1884 and 1890 dealt with this pioneering work, which led him on to the investigation of various types of disruptive discharges in gases, such as lightning, and in 1900 with Hemsalech he showed that when a spark passes through a gas the current is in the first instance carried by the gaseous ions and immediately afterwards by those of the vapourized metal of the electrodes. In 1890 he showed that the experimental work of others on the sparking-distance between electrodes of various forms could be coordinated by calculating the maximum electric field between the electrodes in each case, and in 1904 he determined the coefficient of recombination of the ions in atmospheric air as $3\text{–}4 \times 10^{-6}$ cm.^{3}/sec.

The four solar eclipses in which he had taken observations of the corona led him to consider the question of the constitution of the sun, and its effect on the electrical properties of the earth's atmosphere and the magnetic state of the earth itself. He was one of the first to work out quantitatively the consequences of Balfour Stewart's hypothesis of the existence of a conducting layer in the atmosphere, and he showed in 1889, on the assumption that the forces which produced the daily variation of the magnetic state of the earth had a potential, that those forces resided in the atmosphere and produced electric currents whose effect was modified only slightly by their inductive action on the material below the surface.

In his presidential address to the Physical Society in 1912 he reviewed some of the theories which had been put forward in explanation of the magnetism of the earth and rejected such as were based on electric currents within the surface.

Out of his researches in terrestrial magnetism his general method of searching for hidden periodicities in recorded phenomena, described in his "Periodogram" paper of 1906, appears to have developed.

His early spectroscopic work and his contact with Lord Rayleigh engendered an interest in optical questions which persisted throughout his life. In 1870 his method of adjusting the collimator of a spectroscope appeared, and in 1905 his conclusions as to the best width of slit to use. His simplified methods of dealing with diffraction and interference effects, some based on the hypothesis that light consisted of irregular impulses which were converted into periodic disturbances by the optical apparatus, appeared in a series of papers from 1891 onwards. Absorption and scattering of light (1920) and their bearing on cosmical effects (1905) also interested him.

In other fields he made determinations of the mechanical equivalent of heat, with Gannon, as 4·193 (1895), of electrochemical equivalents, with Crossley (1891 and 1894), of defects of colour vision (1890), and he wrote a series of "electrical notes".

His published works are as follows: 1896, *Intermediate Practical Physics* (with Lees); 1901, *Exercises in Practical Physics* (with Lees); 1904, *An Introduction to the Theory of Optics*; 1911, *The Progress of Physics*, 1875–1908; 1917, *Britain's Heritage of Science* (with Shipley); 1931, *Biographical Fragments*.

The Royal Society awarded him a Royal Medal in 1893, a Rumford Medal in 1926 and a Copley Medal in 1931. He was elected an honorary or a corresponding member of many scientific societies at home and abroad and his fluency in French and German made him invaluable in international conferences. He took a prominent part in the foundation of the International Association of Academies and from 1919 to 1928 was Secretary of the International Research Council. Many foreign delegates to scientific conferences in this country look back with pleasure to his hospitality.

His wide interests in science and his sound common sense rendered him a most desirable member of research committees, and he served for many years on such committees of the Royal Society and the British Association. He had terms of office as Chairman of the Executive Committee of the National Physical Laboratory and as Vice-Chairman of the Meteorological Committee of the Air Ministry. As a Secretary of the Royal Society he took an active part in the organization of scientific research required by government departments during the great war.

He gave much time and thought to the affairs of the scientific and educational institutions with which he was connected and contributed most generously to their finances.

C. H. L.

THOMAS FRANCIS CONNOLLY

Mr Thomas Francis Connolly, who died on November 23, 1934, was born in County Monaghan, Ireland, in 1878. He graduated as M.Sc. at Liverpool in 1899. He spent eight years at the Solar Physics Observatory, Kensington, at the end of which time the Observatory was transferred to Cambridge. Mr Connolly was then appointed Assistant Inspector of Scientific Supplies at the India Store Department. In 1921 he was placed in charge as Inspector of Scientific Supplies. During the war he worked at the Ministry of Munitions in connection with the supply of optical instruments, and was awarded the M.B.E. for his services.

Mr Connolly had a wide range of knowledge, and the secret of his success as an inspector was his capacity for knowing what were the essential parts in the construction of an instrument, and insisting on good workmanship in them. He invented a simple form of theodolite, a reversible level, and a portable standard compass, all three instruments being known by his name. He also published various papers in technical journals dealing with instrument problems.

SIR ALFRED EWING, K.C.B., F.R.S.,

March 27, 1855—January 7, 1935

Many appreciations of the work and personality of Sir Alfred Ewing have appeared in the Press* and to these the reader may be referred who would know something of Sir Alfred as engineer, and something of those physical researches which added so much to our knowledge of the classical magnetism, and gave a new word to the English language. There, too, the reader will find the discreet story of Room 40, and accounts of that genius for administration which revolutionized naval education and guided the destinies of the University of Edinburgh through a critical period of expansion.

Very scant reference, however, has been made to Sir Alfred's work in furthering the study of the scientific basis of refrigeration, and the writer of this note, who has been in touch for many years past with this side of his activities, would place on record a tribute to the memory of one who was a great physicist and a great engineer. Sir Alfred was, indeed, throughout his long working life as much physicist as engineer. His interest in the purely physical side of thermodynamics was as great as his interest in its applications. He possessed a deep knowledge of the work of the founders of the science, and Sadi Carnot's immortal *Réflexions sur la Puissance Motrice du Feu* drew from him unstinted admiration. His own contributions to the science, his *Thermodynamics for Engineers* and his articles in the *Dictionary of Applied Physics* have moulded the thought of some generations of students, and their influence is still profound.

In 1897 he established his reputation as a leader amongst writers on refrigeration by a course of Howard lectures delivered to the Royal Society of Arts. These lectures were published eleven years later under the title *The Mechanical Production of Cold.* The book, charming and lucid in its style, is a classic in the literature of science.

Sir Alfred was the Chairman of a Committee of the Institution of Mechanical Engineers which was set up "to consider and define a standard in refrigeration and its units of measurement". This committee published a report in 1914 in which was given a series of tables and charts of the thermodynamic properties of refrigerants.

When the Food Investigation Board was established, Sir Alfred accepted the chairmanship of its Engineering Committee and continued, without missing a meeting, to preside until his death. He made an admirable chairman, and an investigator of nervous temperament, when presenting his report, had the path smoothed for him by Sir Alfred's genial manner and his clear exposition of the salient points. It was an invariable rule with him to go through the report with its author before the meeting, and any obscurity was cleared up before presentation.

---

* *Times*, January 8; *Engineering*, January 11; *Nature*, January 26, 1935. Sir Alfred's autobiographical volume *An Engineer's Outlook* (Methuen) should also be consulted.

He was keenly interested in the work of the International Union of Physics and took an active part in the work relating to thermodynamic symbols and the definition of the thermal unit. One of his last public appearances was at the meeting of the Union in London in October 1934, when the report of the Symbols, Units and Nomenclature Commission came up for discussion.

His work as Principal of a large university left him, in his later years, small opportunity for research, but on one occasion—when a flowmeter was under consideration—Sir Alfred's enthusiasm was roused to the point of commandeering his bathroom for experiments. An account of the work was published in the *Proceedings of the Royal Society of Edinburgh*. It is interesting to note that this type of flowmeter—a ball in a conical tube—had been suggested by him 50 years previously.

He was a Fellow of the Physical Society for 37 years, and served a term of office as Vice-President.

Sir Alfred Ewing was loved by all who came into intimate contact with him; in losing him the world of science has lost an outstanding personality.

EZER GRIFFITHS

DAVID KING MORRIS

DR DAVID KING MORRIS, a Life Fellow of the Physical Society, passed away after protracted ill-health on May 14, 1935, at the age of 61 years.

Dr Morris became a student in the Engineering Department of University College, London in 1891 under Profs. Hudson Beare, Carey Foster, and J. A. Fleming, and was able to maintain himself at college for the full four-year period by means of the prizes, scholarships and exhibitions which he gained. In the third and fourth years he carried out in conjunction with the present writer research work on dielectric hysteresis and on the measurement of currents in inductive circuits; these were published by the Royal Society and in the *Philosophical Magazine*. He then proceeded as an 1851 Exhibition Scholar to the University and Polytechnicum of Zürich, where he was awarded the degree of Ph.D. in 1897 for a dissertation entitled "The Magnetic Properties and Electrical Resistance of Iron as dependent on Temperature", and on his return to England he was elected a Fellow of University College at the early age of twenty-four.

He was successively demonstrator at the S.W. Polytechnic (London) and Lecturer in Electrical Engineering at the Mason University College (which later became the University) at Birmingham. In the latter position Dr Morris was responsible for the entire work of the Electrical Engineering Department and in addition to his teaching duties much valuable research work was carried out or inspired by him.

The equipment of the new laboratories, admittedly at the time one of the finest in the country, was undoubtedly the result of his experience and foresight.

In the years 1904–1908 he contributed four joint papers to the Institution of Electrical Engineers. The paper on eddy-current brakes for testing motors (1905) was in front of its time, and the mathematical theory developed by him is now found to be of real service in engineering practice in the design of brakes for electricity-meters.

He later helped to found the firm of Morris and Lister at Coventry.

During the war as director of the M.L. Magneto Syndicate he became one of the chief suppliers of magnetos for aircraft.

A. W. P.

ARTHUR LYNCH

Colonel Arthur Lynch, who was born at Smythesdale, near Ballarat, in Australia in 1861 and died in a London nursing home this year, was almost a legendary figure. To some who were permitted to glimpse his achievements in the most diversified fields he may well have appeared something like a universal genius, and that is an interpretation of his character which he would not have put himself to any particular pains to contradict. Others, meeting him in his more expansive moods, may have been inclined to murmur "charlatan", in which they would have done him less than justice.

To the public in the opening years of the century Colonel Lynch was "the man who fought for the Boers" and was sentenced to death and reprieved. Opinions differ about the importance of the Irish Brigade which he is understood to have commanded, but it may be said with some confidence that it made no great difference to the military history of the campaign, and it was a wise act of clemency that made it possible for the rebel "Col." Lynch of the Boer war to be colonel of the 10th Battalion, Munster Fusiliers in France in 1918. From 1901 to 1903 he sat in the House of Commons as Nationalist member for Galway and he returned to the House in 1909 and sat until 1918 for West Clare. When the propriety of putting him forward in 1909 was challenged by a timorous member of the Selection Committee on the ground that he had "killed Englishmen", another member with an acid wit which Lynch himself could appreciate, retorted that the Colonel had killed more Englishmen when he was a doctor in London than when he was a soldier in South Africa.

That was not a verdict he would have accepted, for he took himself seriously as a scientist. "I have had", he used to claim, "a longer academic training than any other man living." After leaving Melbourne University with degrees in arts and engineering, he studied at Berlin University, took the diploma of the École Supérieure d'Électricité in Paris, followed a medical course at the Hôpital Beaujon in Paris and qualified for the Medical Register in this country by taking the conjoint examination from St Mary's Hospital, entitling him to inscribe himself M.R.C.S., L.R.C.P.

His engineering studies were interrupted by the attractions of journalism in Paris, and it is probably a sound estimate of his career to say that he was essentially a journalist. He had that convenient type of mind that enables its possessor to pass any examination with little effort, and a real flair for "news value". To this must be added an Irishman's love of being against the conventions. He was a warm supporter of Spahlinger's claims in the field of tuberculosis, but a caustic and unsparing critic of the psycho-analysts.

His attitude towards scientific questions is embodied in a pungent volume which he wished to call *Humbugs of Science*. The late Sir John Murray, however, preferred the more decorous title *Science, Leading and Misleading*, under which it ultimately appeared. But the fight which he considered "the best and the last" was against Einstein. He regarded relativity physics as a confused sort of metaphysics and ultimately got his book published. His confident declarations to publishers that he could "knock Einstein out in one round" delayed the issue.

Lynch published novels, poems, a two-volume book on psychology which, he claimed, laid the foundations of a new and true system, and a book on the future of the Empire. The book he most wanted to write at the end was a defence of republicanism, in which he held a passionate belief that not even first-hand experience of republics had succeeded in eradicating.

REVIEWS

Electron Emission and Adsorption Phenomena, by J. H. de Boer. Translated from MS. by Mrs H. E. Teves-Acly. (Cambridge: The University Press.) 21*s.*

The author of this book is on the research staff of the Philips Lamp Factory at Eindhoven. It contains an account of thermionic and photoelectric effects, both external and internal, written with special reference to the influence of adsorption at the outer surfaces and internal flaws of the material. Following a preliminary account of thermionic emission, the second chapter gives a careful discussion of the forces which can give rise to adsorption, and their relative importance in the case of atoms and ions, polar and non-polar systems. Apart from a digression in chapter vii to give an account of the photo-ionization of gaseous molecules and the Franck-Condon principle, the remainder of the book shows how the ideas of the second chapter can be applied first to the emission of electrons from complex surfaces such as are widely used for dull-emitter filaments, and next to the phenomena of internal conduction associated with the names of Gudden and Pohl.

A very large number of experimental results are described, including many of the author's own; these, however, seem to have been selected with a view to illustrating the theory rather than to giving a complete account of the subject. For example, in the chapter on the selective photoelectric effect only two lines deal with the liquid sodium-potassium alloy with which so much of the work has been done. Again, more reference might have been made to the work of Przibram and his school on the coloration of alkali halides.

The author's views, which are a development of ideas due to Langmuir, Becker and others, are well set out. On a basis of the forces between surfaces and the atoms adsorbed on them he is able to explain a wide range of phenomena, many of them of commercial importance. Objections are dealt with fairly, and the general impression with which one is left is that the author has succeeded in presenting a probable explanation of the main facts and one capable of extension in other directions.

The style is, in some places, decidedly clumsy, perhaps as a result of too literal a translation, but in spite of this the book is readable.

G. P. T.

The Optical Basis of the Theory of Valency, by R. de L. Kronig, Ph.D. Pp. x + 246. The Cambridge Series of Physical Chemistry. (Cambridge: The University Press, 1935.) 16*s.* net.

Knowledge of the nature of chemical valency has developed rapidly in recent years from G. N. Lewis's hypothesis of octets and sharing of paired electrons to the more detailed theories of Heitler and London (1927), Pauling and Slater (1931) and Hund and Mulliken (1932–3). This development is a consequence of the progress made in the quantum-theoretical interpretation of molecular spectra, and has been achieved very largely by the application of purely physical methods. It is not only the chemist who has found difficulty in keeping abreast of this outgrowth of modern physics and felt the need of a clear exposition in English book form. To present the subject adequately within the rather narrow limits of some 250 pages is no easy task for any author, be he theoretical physicist, physical chemist or spectroscopist. However, it has been achieved most admirably in this work by Dr Kronig, who has already won world-wide fame not only by his important contributions to modern theoretical physics but also by his powers of lucid exposition of the essentials of the subject in our language as well as his own.

Taking advantage of the fact that detailed accounts of the analysis of atomic and molecular spectra and of the quantum-mechanical theory of spectra and valency are available in other recent books, Dr Kronig has reduced to a minimum both the mathematical argument and the spectroscopic detail, thereby widening the scope of the work and making it more acceptable to the chemical reader. As will be seen from the following outline it is not entirely spectroscopic, nor is it for the chemist only. The opening chapter deals with relevant chemical facts, atomic theory, atomic models, etc., the next with the investigations of atomic and molecular structure by means of X-ray and electron diffraction. The third chapter is given to atomic spectra and the periodic system, and the next two to band spectra and chemical binding in diatomic and in polyatomic molecules. The phenomena of optical and thermal dissociation are dealt with in the last chapter. A table of atomic electron configurations and ground states is appended.

Some of the notes made in the course of a first reading may not be out of place here. (i) It is clear from the first paragraph of p. 88 that α in equation (21·15) is a positive quantity; hence equations (21·16) and (21·20) must, if they are to agree with observation, have + in place of their − signs, whatever the theoretical justification for the latter may be. (ii) It is no longer true of all the measurements of rotation bands that, as is stated on p. 95, they were made in absorption and so refer to $v=0$. Actually Strong has measured fairly extensive rotation bands of HCl in emission for $v=0$ (to $J=33$) and for $v=1$ (to $J=23$). (iii) The table of constants for the ground states of diatomic molecules on pp. 100–4, in which the author has supplemented the data in this Society's report by more recent data (to April 1935), is by no means complete; many emitters recorded before that date are omitted, and for a few others the data are not the latest. NiH looks ill at ease in the class of halogen hydrides on p. 101. (iv) The proximity of the dissociation energies of N_2 and CO is closer than appears on p. 118, the latest value of the latter being as low as 8·41 V. as against the earlier estimate of 10 V. quoted there and on p. 136. (v) The spectrograms in the plates facing pp. 170 and 222 would gain in usefulness if provided with scales of either wave-lengths or wave-numbers. (vi) In the 12th line from the foot of p. 236 either the two ω_e values should be interchanged or the word "respectively" should be deleted.

This is an important work, which should find an even wider circle of readers than the author's well-known *Band Spectra and Molecular Structure*. As usual, the work of the Cambridge University Press is beyond praise.

W. J.

La Spectroscopie appliquée, by P. Swings. Pp. 188. (Bibliothèque scientifique Belge. Paris: Hermann et C^ie^, 1935.) 15 francs.

In spite of the great progress made in spectrographic analysis in recent years, it appears that the method has not been applied in Belgium to nearly the same extent as it has in many other European countries and the U.S.A. To remedy in some measure this unsatisfactory state of things is the main object of this excellent little monograph, which is based to a large extent on a course given by the author in the Technical Faculty of the University of Liège, and is the first of its kind in the French language. An entirely practical point of view is adopted throughout. After an introductory chapter on the history, advantages and physical basis of the spectroscopic method of analysis, we have one chapter on instruments (spectrographs and light-sources), two chapters on qualitative and quantitative analysis by means of emission spectra, and two on typical applications to chemical, technical and biological problems. Then follow a chapter on analysis by means of absorption spectra, one on other methods (infra-red, X-rays and fluorescence), and one outlining the physical interpretation of atomic and molecular spectra. A useful bibliography and a folding plate of spectrograms are appended. This small book contains a remarkable amount of useful information, which is both well-selected and well arranged. It can be thoroughly recommended to English readers.

W. J.

Molekülspektren und ihre Anwendung auf Chemische Probleme: I, Tabellen, by H. SPONER. Pp. vi + 154. (Berlin: Julius Springer, 1935.) RM. 16, bound RM. 17.60.

The usual plan of tabulating numerical data in an appendix to the main text of a comprehensive work, especially one on such a wide and rapidly growing subject as molecular spectroscopy, has the disadvantage that it makes no provision for two probable requirements: many readers may require the tables only, and the tables may require revision and expansion more frequently than the text. Prof. Sponer and her publishers have therefore done wisely in issuing these valuable tables as a separate volume, which is to be followed by the text in volume II. The work is an addition to the well-known series of *Struktur und Eigenschaften der Materie* monographs founded by Professors Born and Franck.

Both diatomic and polyatomic molecules are included, each class occupying about a half of the volume. For diatomic molecules there are tables of constants for all the known electronic states and band systems, observed predissociation effects, isotope effects, alternating intensities and nuclear spins, and electron-impact processes. Most of these are founded, as Dr Sponer very handsomely acknowledges, upon the corresponding tables (Appendix II and Tables 12 and 26) of this Society's Report on the subject, supplemented where possible by data as to the electron configurations and dissociation products of the various molecular electronic states, as well as by new data from original papers that have appeared since 1932. Those chemists and spectroscopists who have pointed out the need of a new edition of Appendix II will certainly find here all they require for the time being. The tables for polyatomic molecules are entirely new; they cover infra-red absorption and Raman spectra (with diagrams of the modes of vibration of the various molecules in their ground states), electronic bands, and electron-impact processes. The work is brought up to date as far as possible by two supplementary tables containing matter which was published in 1934 and early 1935 while the rest of the volume was going through the press. Each group of tables is preceded by a few explanatory paragraphs and followed by a list of references to original papers, and an adequate eight-page index is provided.

Great care has obviously been taken in the preparation of the tables, and the omissions and numerical slips so far detected are very few indeed. Perhaps it would be well to point out here that the entries 6·266 and 5·181 on p. 13 for two states of Li_2 are the values of $2x_e\omega_e$, not $x_e\omega_e$ itself; this error is the result of a rather unusual method of presentation in the original paper. Spectroscopists and physical chemists will undoubtedly find these tables invaluable and will look forward to the appearance of the text itself.

W. J.

Optical Rotatory Power, by T. M. LOWRY, C.B.E., F.R.S. Pp. xiii + 483. (London, Longmans, Green and Co., Ltd.) 30*s*. net.

This is, we believe, the first original English work on a subject which, during the century or more that has passed since Biot's discovery, has attracted much attention, and has accumulated a wealth of special literature. Prof. Lowry's volume is hardly likely to find a rival in our generation. His treatment of the subject which he has made so peculiarly his own is scholarly, full and well-balanced, and the citations from the literature are detailed enough to save even the specialist much rummaging in the dusty files of the original papers.

The scope of the book may be indicated sufficiently by the titles of the four parts into which it is divided—*Historical and General*, *Polarimetry*, *Special Cases*, and *Theoretical Considerations*. Theory and experiment are held admirably in balance, the book is well up to date—Born's latest views receive adequate and detailed treatment—and the descriptions of apparatus and instruments leave nothing to be desired. The book, indispensable to the physical chemist, may be read with pleasure and profit by anyone interested in the history and advance of purely physical science.

A. F.

The Theory of Atomic Spectra, by E. U. CONDON, Ph.D., and G. H. SHORTLEY, Ph.D. Pp. xiv + 441. (Cambridge: The University Press, 1935.) 42*s*. net.

The theoretical interpretation of the structure of line spectra of neutral and ionized atoms has now extended over a little more than two decades. In the first Bohr's quantum theory, a combination of quantum postulates with classical kinematics and dynamics, was developed, and its limitations were ascertained. In the second the new dynamical theory was developed along two different lines, the matrix mechanics by Heisenberg and the wave mechanics by Schrödinger, afterwards unified by Dirac to form the system of quantum mechanics, in terms of which all the observed structural features of atomic spectra are much more completely understood and at least semi-quantitatively explained. In addition to the many books of various degrees of difficulty which have appeared on quantum mechanics itself, we have had several on the theoretical interpretation of atomic spectra, which, with one notable exception, were written from the point of view of Bohr's correspondence principle. The exception is White's *Introduction to Atomic Spectra* (McGraw-Hill, 1934) which meets the requirements of those taking up the subject for study in the laboratory, especially if it is used in conjunction with Bacher and Goudsmit's *Atomic Energy States* (McGraw-Hill, 1932). Although several fundamental theoretical problems still await solution, and more detailed theoretical calculations on atomic spectra are desirable, the present state of the interpretation of such spectra may be regarded as fairly closed and highly satisfactory, and, therefore, a fitting subject for a complete and authoritative survey such as is provided in this masterly treatise.

In the introductory chapter the authors give a brief and well-balanced historical sketch, and in the second an exposition, along the lines laid down by Dirac, of the quantum-mechanical principles used in the rest of the work. The next two chapters are devoted to the presentation of special results of the general theory which are used throughout the book. The detailed development of the theory of atomic spectra begins with chapter v and continues as a unified and logical deduction of their structure from quantum mechanical principles throughout the remaining fourteen chapters. The theory is given in complete detail and the results are compared with experimental facts, but no attempt is made to adhere to chronological order. While admitting the value of the group theory in atomic physics the authors derive all the results without it in order to avoid unnecessary addition to the mathematical apparatus; there are, indeed, already three recent books by Weyl, Wigner and van der Waerden in which atomic spectral theory is regarded from the standpoint of the group theory. A discussion of universal constants and natural atomic units is appended. The book is well indexed, and, as goes without saying, magnificently produced. To describe it as a monumental work is no idle repetition of a reviewer's stock phrase; it really is one. It will be especially welcome to those spectroscopists who have for some time past had to seek theoretical results in widely scattered papers, as well as by those who wish to continue their reading beyond White's book, which, indeed, will serve as an excellent introduction to this fuller work.

W. J.

A Text-book of Light, by L. R. MIDDLETON. Pp. viii + 288. (London: G. Bell and Sons, Ltd.) 6*s*.

This book covers the ground required for the subject of Light in the Higher School Certificate and the Intermediate Science examinations, although it aims at a treatment rather more general than that of a text intended only for preparation for such tests.

It may be said at once that the author has a clear style of exposition. The general scope of the book is well conceived, and an attempt is made to awaken interest in optical instruments and the phenomena of physical optics. It is disappointing, however, to find that he has completely ignored the recommendations of the Physical Society's *Report on the Teaching of Geometrical Optics*, so that the use of the book can only tend to perpetuate the unhappy confusion and division already existent between school optics on the one hand and technical optics on the other. Furthermore, the most objectionable features of the usual school optics, such as the use of unsymmetrical equations of the type

$$\frac{\mu}{v}-\frac{1}{u}=\frac{\mu-1}{r}$$

are again encountered.

While a desire to awaken an interest in optical instruments is praiseworthy in an author, the appearance of such general statements regarding a telescope as

$$\text{“magnifying power} = \frac{\text{diameter of objective}}{\text{diameter of eyepiece}}\text{”}$$

(p. 123) may very justly disturb our confidence in the author's wisdom in venturing off the beaten track. The paragraph on "night glasses" is really entirely concerned with explaining the enhanced visibility of stars obtained with a telescope, and has little or nothing to do with "night glasses" in the ordinary sense of the term. The whole section is most misleading, and the same thing must be said of some other parts of the book.

Much better things than this must be expected before there can be any valid reason for abandoning some of the older school books of the same type. L. C. M.

Origins and Development of Applied Chemistry, by Prof. J. R. PARTINGTON, M.B.E., D.Sc. Pp. xii + 597. (London: Longmans, Green and Co., 1935.) 45*s*. net.

In spite of all that has been written on the subject during the last thousand years, the plain truth is that the origin of chemistry remains an unsolved mystery. The course of the science from Islam to Latin Christianity, and its subsequent growth to the majestic proportions that characterize it to-day, are clear enough. It is also pretty well established that the Muslims obtained their initiation into chemical knowledge mainly by way of Alexandria, and that the Alexandrian Greeks derived much from Egyptian sources. But the simple hypothesis that chemistry arose *ab initio* in ancient Egypt, and that the chemical succession is unbroken through Moses, Mary the Copt, Zosimos the Panopolitan, Khalid, Jabir, Razi, Roger Bacon, Albert the Great, Paracelsus, Lémery and Boyle, has little but its simplicity to recommend it. We now know, for example, that Assyria possessed competent practical chemists as early as the seventh and sixth centuries B.C.; that alchemy in China is at least as ancient as in Alexandria, and possibly much older; that much the same is true of chemical theory and practice in India; and that Persian alchemy shows many signs of a very remote indigenousness. The wealth of discoveries of this kind published during the present century has thus not made the problem of the origin of chemistry any easier; on the contrary, it has shown that the whole matter is far more complicated than had been suspected, and that a great deal of research will have to be carried out before any sort of finality is reached. It has also become clear that the data for speculation are insufficient, and consequently require to be supplemented by the patient accumulation of as many relevant facts as possible.

This kind of work is exacting, tedious and pedestrian. It offers no alluring prospect of spectacular discoveries, and demands the closest attention and perseverance in the establishment of detail. It is, in short, the kind of work that everybody would like to see done, but hopes that someone else will do. Prof. Partington has unselfishly attacked the task, and all who read the present book will agree that he has made a noble beginning. He is, of course, admirably equipped for the purpose, most especially in his unswerving determination to be accurate, to give line and word of his authorities, and to refrain from supposition; with Newton he would say *hypotheses non fingo*. He has, however, the additional qualification of possessing as wide a knowledge of old chemical literature as any man now living, with the probable exception of Prof. E. O. von Lippmann—and von Lippmann read the greater part of the book while it was still in manuscript. There could thus have been no more appropriate a choice for the work, and we must count ourselves fortunate that Prof. Partington, rejecting fields in the history of chemistry where honour is more easily to be gained, has methodically dug ground too intractable and too arid for any of less resolution than he.

The plan of the book is simple. Prof. Partington deals in turn with Egypt, Babylonia and Assyria, the Aegean Civilization, Troy and Cyprus, Asia Minor, Persia, Phoenicia, and Palestine. In each part, a short account of the general history and archaeology of the region concerned is given first (an extremely useful feature which the reader will appreciate) and this is followed by sections on metals and non-metals; each section is further divided into sub-sections, within which individual metals, minerals and other substances are separately considered. By this arrangement it is easy for the reader either to get a view of the technical chemical knowledge of a particular region, or to turn up all that was known of a given substance in the whole of the various regions.

It is difficult in a review to indicate the wealth of facts which Prof. Partington has amassed, docketed, arranged and fully indexed. There are, for instance, over 100 index references to lead and about the same to iron, while gold, as might be expected, occupies more than half a column of the subject index. Within the limits he has chosen, it would be hard to find any fact of importance that Prof. Partington has omitted; but not content with mere compilation, he has added comments and corrections from the chemist's point of view in cases where the facts reported are obviously doubtful or erroneous.

The general impression left by the book is one of surprise at the wide, deep and accurate knowledge of an empirical chemical kind possessed by the nations of antiquity. It seems, also, as if no one nation showed any remarkable pre-eminence over the others, with the possible exception of Egypt; and it is thus not very hazardous to suggest that research will have to be pushed back still further before the period and region of the principal discoveries—e.g. the smelting of ores—can be definitely fixed. Furthermore, the problem of the origin of the corpus of alchemical doctrine does not yet make contact with the chemical facts recorded by Prof. Partington; but his book is a long step towards the provision of a *terminus a quo* in this connexion.

Those who know the discipline of historical research in science will best appreciate Prof. Partington's achievement; but it would be a mistake to look upon his book as one for the specialist only. The general reader interested in the story of man's gradual conquest of the material world will find it a storehouse of relevant and reliable information, presented in a scholarly and attractive manner.

E. J. HOLMYARD.

A Symposium on Illumination, edited by WEBBER GRIEVESON, B.Sc., M.A., with a foreword by Lt.-Col. K. EDGCUMBE. Pp. xv + 229. (London: Chapman and Hall, Ltd.) 13*s*. 6*d*. net.

In 1933 a series of lectures was given (mainly at the Regent Street Polytechnic) under the auspices of the National Illumination Committee and the Illuminating Engineering

Society. The lectures, which are reprinted in this volume, demonstrate admirably the amazingly rapid advances made in this, as in so many other branches of physics, pure and applied. The topics considered are best shown by a catalogue of the lecture titles: Lighting in the Service of Mankind (C. C. Paterson); Radiation—Light and the Eye (J. W. T. Walsh); Electric Lamps and their Characteristics (W. J. Jones); Gas Lamps and their Characteristics (F. C. Smith); Daylight: Coloured Light (J. W. T. Walsh); Photometric Standards, Lamp Photometry and Illumination Measurement (J. T. MacGregor-Morris); The Redistribution of Light (G. H. Wilson); Lighting for Decoration and Entertainment (A. B. Read); Public Lighting (J. F. Colquhoun); and Lighting for Safety, Health and Welfare (H. C. Weston).

The treatment is what has been called semi-technical; it is popular in the best sense of the word; it is never cheap; and the lectures hold the reader's attention from the opening sentence of Mr Paterson's brilliant discourse, to that last page on which Mr Weston gives sage advice on spectacles for fine work.

A book to be thoroughly commended.

The Scientific Journal of the Royal College of Science. Vol. v. Pp. 137. (London: Edward Arnold and Co.) 7*s.* 6*d.* net.

The topics dealt with in the fifth volume of this journal are perhaps a little more pedestrian than those discussed in previous issues. Nevertheless the volume maintains the high standard of its predecessors, and contains adequate reports of some very valuable lectures. Seven addresses are reported under the heading of the Chemical Society's section, four under the Natural History section, and three only under the section allotted to the Mathematical and Physical Society, of which one address of special interest—that of Prof. Milne on the Expanding Universe—is reported only in abstract. We trust that succeeding issues of this journal, the importance of which increases with each year of issue, will see a more even distribution of topics between the sections allotted to the three Societies concerned.

A. F.

Actualités Scientifiques et Industrielles. (Hermann et Cie., 6 Rue de la Sorbonne, 6, Paris.)

Under the above general heading we have received the monographs listed below. Each is written by an authority on his subject and the treatment is, in general, concise and clear.

222. Y. Rocard. *Propagation et Absorption du Son.* 15 fr.
225. H. Mineur. *Dénombrements d'Étoiles.* 15 fr.
230. P. M. S. Blackett. *La Radiation Cosmique*, I. *Aperçu Général.* 10 fr.
231. P. M. S. Blackett. *La Radiation Cosmique*, II. *La Méthode de la Chambre de C. T. R. Wilson.* 8 fr.
232. P. M. S. Blackett. *La Radiation Cosmique*, III. *L'Action du Champ Magnétique Terrestre.* 7 fr.
233. P. M. S. Blackett. *La Radiation Cosmique*, IV. *La Perte d'Energie par Ionisation.* 10 fr.
234. Y. Rocard. *La Stabilité de Route des Locomotives.* 15 fr.
236. Pierre Fleury. *Généralités sur les Mesures.* 15 fr.
241. P. Swings. *Les Spectres des Nébuleuses Gazeuses.* 10 fr.
245. Manuel Valadares. *Transmutation des Eléménts.* 10 fr.
247. G. Guében. *Structure Nucléaire.* 10 fr.
248. B. Rossi. *Rayons Cosmiques.* 12 fr.

X-Rays in Theory and Experiment, by A. H. COMPTON and S. K. ALLISON. Pp. xiv + 828. (London: Macmillan and Co., 1935.) Price 32*s*. 6*d*.

Compton's *X-Rays and Electrons*, published in 1926, was a book of very strongly marked individuality; it was, moreover, largely given up to the study of a relatively small number of the outstanding X-ray problems of its time. This being so, it is more than usually interesting to see how the book transforms, nine years later, into a second edition.

In the first place, there has been an inevitable increase in bulk; the new work is more than twice the size of its predecessor. Secondly, Compton and Allison have transformed the book into a systematic and just sufficiently encyclopaedic treatise which covers a considerable portion of the whole domain of the physics of X-rays. In brief, the change in the book is well enough indicated by the change in the title; the new edition justifies the new title, comprehensive though that is.

The chapter headings give the clearest indication of the scope of the work. They are: (1) Discovery and Properties of X-rays (55 pages); (2) Production of X-rays (60 pp.); (3) Scattering of X-rays (147 pp.); (4) Dispersion Theory applied to X-rays (53 pp.); (5) Study of Crystal Structure by means of X-rays (49 pp.); (6) Intensity of Reflection of X-rays from Crystals (107 pp.); (7) Phenomena associated with the Ejection of Photo-electrons by X-rays (111 pp.); (8) Interpretation of X-ray Spectra (89 pp.); (9) Some accurate Methods of X-ray Wave-length Measurement and their Results (85 pp.). There are eleven Appendices of very varied content, which includes some valuable absorption data, as well as matter of more theoretical interest.

In general most stress is laid upon those divisions of X-ray research which are most germane to the major problems of atomic structure and of radiation. Other sections—on which one would normally consult such treatises as those of the Braggs or Siegbahn—are developed as a rule in less detail. This is not, however, to say that they are inadequately handled: in fact their salient features are in general depicted in a most illuminating fashion. The result is that while the book is still very directly addressed to the X-ray specialist, it is also fitted to serve as a scholarly and inspiring introduction to the subject for readers whose interests are less localized. There is an extensive bibliography for which completeness is not claimed, but which is ample for the needs of those beginning serious research in the physics of X-rays.

It is pleasant to note that in the course of its revision and expansion the book has not lost its original air of spontaneity; it still smells healthily of the laboratory. In other words the authors, having played an active part in building up the experimental material which they discuss, have naturally succeeded in presenting that material as part of a live and vigorous organism.

This is a notable and much-needed addition to the literature of X-rays, and we are fortunate in more ways than one in having it presented to us in English. It is salutary to think of the price which would have been demanded for a treatise of similar extent if it had happened to bear the title "Theoretisches und Experimentelles über Röntgenstrahlen". As it is, we have a great deal of good book for a relatively very modest price.

Printing and production are excellent, but it would assist cross-reference (especially in view of the way in which the equations are numbered) if the chapter numbers were inserted at the head of each page. For this, the present reviewer would willingly sacrifice the subject titles which head alternate pages, and which must have been far more troublesome to draft and print.

H. R. R.

INDEX

PAGE

Absorption factor for the powder and rotating-crystal method of X-ray crystal analysis, The 879
Acetylene flames, The colour temperatures of the Hefner and 1032
Adsorbed upon thermionic, photoelectric and catalytic surfaces, Models of the superposition and interpenetration of components in gas mixtures . . . 287
Adsorption of gases on mercury 460
Aerial revolving in its own plane, The current variations in a short-wave square frame 388
Aerials, On a theory of the action of rectangular short-wave frame 377
Aluminium, The longitudinal thermoelectric effect 859, 862
Appleton, E. V. and D. B. Boohariwalla: The influence of a magnetic field on the high-frequency conductivity of an ionized medium 1074
Argon, The energy of agitation of positive ions in 74
Astbury, N. F.: The computation of the integrals required in mutual-inductance calculations 86
Asundi, R. H., R. Samuel and M. Zaki-uddin: The band systems of cadmium fluoride 235
Atmosphere, Ionization charts of the upper 263
Awbery, J. H. and Ezer Griffiths: An investigation of the wet-and-dry-bulb hygrometer at low temperatures 684

Balls, E. G., *see* Nettleton
Band spectrum of beryllium monoxide, The 415
Band spectrum of vanadium oxide, The 433
Band systems of cadmium fluoride, The 235
Band, William: The longitudinal thermoelectric effect: (4) A further study of aluminium 862
Band, William, *see also* Ch'en, Hsü, Li, and Pi
Bands by electron impact, The efficiency of excitation of the nitrogen first positive 420
Bands of phosphorus monoxide, Rotational analysis of the ultra-violet . . . 247
Barasch, H. P.: An improved counting circuit 824
Barnard, G. P.: A new selenium-sulphur rectifier photoelectric cell 477
Bates, L. F. and D. V. Reddi Pantulu: The magnetic properties of amorphous manganese 197
Bates, L. F. and H. E. Hogwood: A note on the Raman spectrum of a ferromagnetic oxide 877
Bell, H., *see* Harvey
Bentham, W. E.: Electronic theory and the magnetron oscillator 1
Beryllium monoxide, The band spectrum of 415
Bisacre, F. F. P.: Convergent polarized light and Hertz's problem for a uniaxial material 306
Bisacre, F. F. P.: The theory of the formation of an image by a plane band grating used in the soft X-ray region 948
Bond, W. N.: The surface tension of a moving water sheet 549
Boohariwalla, D. B., *see* Appleton
Boron by neutrons, The disintegration of 873
Bradley, A. J.: The absorption factor for the powder and rotating-crystal methods of X-ray crystal analysis 879
Brentano, J. C. M.: The quantitative measurement of the intensity of X-ray reflections from crystalline powders 932

PAGE

Brown, G. Burniston: On vortex motion in gaseous jets and the origin of their sensitivity to sound . . . 703
Builder, Geoffrey and A. L. Green: Modulation-frequency-change technique for ionospheric measurements . . . 1085
Bunn, C. W.: The lattice-dimensions of zinc oxide . . . 835
Burdon, R. S.: Adsorption of gases on mercury . . . 460

Cadmium fluoride, The band systems of . . . 235
Campbell, N. R.: The statistical theory of errors . . . 800
Campbell, N. R. and G. C. Marris: The measurement of loudness . . . 153
Carbon, Structure of the K line of . . . 626
Cerium chloride in aqueous solution, Magnetic susceptibility of . . . 559
Chalmers, Bruce: An interference extensometer and some observations on the elasticity of lead . . . 352
Chalmers, Bruce: The twinning of single crystals of tin . . . 733
Ch'en, J. L. and William Band: Longitudinal thermoelectric effect: (5) Silver . . . 904
Cherry, R. O., *see* Martyn, D. F.
Clews, C. J. B.: The electrical conductivity of some strong electrolytes in dilute solution and its variation over the temperature range 18° C. to 85° C. . . . 818
Chipman, R. A.: The electron-oscillation characteristics of an experimental plane-electrode triode . . . 1042
Cockburn, R.: The variation of voltage-distribution and of electron transit-time with current in the planar diode . . . 801
Colorimeter, A new precision . . . 69
Colorimeter, A trichromatic . . . 1068
Colorimetry, Rapid mathematical methods for trichromatic . . . 400
Colour-matching of tungsten-filament lamps by means of a single photocell and colour filters, The . . . 1012
Colour temperatures of the Hefner and acetylene flames, The . . . 1032
Colour-vision characteristics of two trichromats, The . . . 205
Compton, Arthur H.: The twentieth Guthrie Lecture: An attempt to analyse cosmic rays . . . 747
Conductivity of an ionized medium, The influence of a magnetic field on the high-frequency . . . 1074
Conductivity of some strong electrolytes in dilute solution and its variation over the temperature range 18° C. to 85° C. . . . 818
Connolly, T. F., Obituary notice of . . . 1134
Cooke, A. H., *see* Simon, F.
Cosmic rays, An attempt to analyse . . . 747
Counting circuit, An improved . . . 824
Crater, Candle-power per ampere of the positive . . . 667
Crystal analysis, The absorption factor for the powder and rotating-crystal methods of X-ray . . . 879
Crystalline powders, The quantitative measurement of the intensity of X-ray reflections from . . . 932
Crystals of tin, The twinning of single . . . 733

Dent, Beryl M.: On observations of points connected by a linear relation . . . 92
Deuterium by electrolysis, The efficiency of separation of hydrogen and . . . 1060
Dielectrics under high voltage, The breakdown of . . . 974
Diffraction grating, The use of a liquid surface carrying ripples as a . . . 998
Dines, L. H. G.: The potential acquired in the natural electric field by a vertical rod 218

PAGE
Diode, The variation of voltage-distribution and of electron transit-time with current in the planar 801
Discharge tubes, Spectroscopic observations of recurrent phenomena in . . 1019
Donaldson, R.: A trichromatic colorimeter 1068

Earth and the measurement of the resistivity of the earth, Experiments relating to the distribution of alternating electric currents in the 589
Electrolytes in dilute solution, The electrical conductivity of some strong . . 818
Electrolytes, The temperature variation of the viscosity of aqueous solutions of strong 657
Electron diffraction, Oxide films on liquid metals studied by means of . . . 109
Electron-oscillation characteristics of a plane-electrode triode 1042
Electronic oscillations, Some experiments on 277
Electronic theory and the magnetron oscillator 1
Errors, The statistical theory of 800
Ewing, Sir Alfred, Obituary notice of 1135
Extensometer and some observations on the elasticity of lead, An interference . 352
Eyles, E. D. and E. W. H. Selwyn: New method of measuring the time and efficiency of photographic shutters, A 446

Films on liquid metals studied by means of electron diffraction, Oxide . . 109
Follett, D. H.: The use of microphotometric methods in divided-beam spectrophotometry 125

Gas mixtures adsorbed upon thermionic, photoelectric and catalytic surfaces . 287
Grating used in the soft X-ray region, The theory of the formation of an image by a plane band 948
Green, A. L., *see* Builder, G. and Martyn, D. F.
Griffiths, Ezer, *see* Awbery, J. H.
Gupta, A. K. Sen: Rotational analysis of the ultra-violet bands of phosphorus monoxide 247
Guthrie Lecture, The twentieth 747

Harmonic analysis, A negative-resistance device and its application to . . 471
Harvey, A. and H. Bell: The band spectrum of beryllium monoxide 415
Heat, The effect of the aeration of the water used in the determination of the mechanical equivalent of 1003
Hefner and acetylene flames, The colour temperatures of the 1032
Hercus, Associate-Professor E. O., *see* Laby, Prof. T. H.
Hertz's problem for a uniaxial material, Convergent polarized light and . . 306
Hogwood, H. E., *see* Bates, L. F.
Hollingsworth, Professor J.: The structure of the ionosphere 843
Holmes, J. G.: Rapid mathematical methods for trichromatic colorimetry . . 400
Hsü, Y. K. and William Band: Thermomagnetic hysteresis in nickel wire . . 910
Hydrogen and deuterium by electrolysis, The efficiency of separation of . . 1060
Hydrogen by the expansion method, Liquefaction of 678
Hygrometer at low temperatures, An investigation of the wet-and-dry-bulb . 684
Hysteresis in nickel wire, Thermomagnetic 910

Inductance and effective resistance of iron-cored chokes carrying direct current, A dynamometer null method of measuring the 964
Ionization charts of the upper atmosphere 263
Ionized medium, The influence of a magnetic field on the high-frequency conductivity of an 1074

PAGE

Ionosphere, The propagation of medium radio waves in the 323
Ionosphere, The structure of the 843
Ionospheric measurements, Modulation-frequency-change technique for 1085
Ions in argon, The energy of agitation of 74

Jackson, L. C.: The magnetic moment of the manganic ion 1029
Jeffreys, Harold: Time and amplitude relations in seismology 455
Jenkins, R. O.: Oxide films on liquid metals studied by means of electron-diffraction 109
Johnson, M. C.: Models of the superposition and interpenetration of components in gas mixtures adsorbed upon thermionic, photoelectric and catalytic surfaces 287
Jones, D. E. H., *see* Macgregor-Morris, J. T.
Jones, F. Llewellyn: The energy of agitation of positive ions in argon 74
Jones, H. Goulbourne: Measurements of the thermal expansion of cast and rolled zinc 1117

Kemp, C. F. B., Obituary notice of 1129

Laby, Professor T. H. and Associate-Professor E. O. Hercus: The effect of the aeration of the water used in the determination of the mechanical equivalent of heat 1003
Lead, An interference extensometer and some observations on the elasticity of 352
Lee, H. W., *see* Taylor, W.
Lens, The development of the photographic 502
Leyshon, W. A.: Some experiments on electronic oscillations 277
Li, M. K. and William Band: The longitudinal thermoelectric effect: (3) Aluminium 859
Linear relation, On observations of points connected by a 92
Liquefaction of hydrogen by the expansion method 678
Liquier-Milward, Jeanne: Magnetic susceptibility of cerium chloride in aqueous solution and its variation with temperature 559
Loudness, The measurement of 153
Lynch, A., Obituary notice of 1137

Macgregor-Morris, Professor J. T. and D. E. H. Jones: A spectrographic examination of arcs between plain soot carbons and its connection with the candle-power per ampere of the positive crater 667
Maddock, A. J.: Absolute intensities in the spectrum of quartz mercury arcs and their variation with temperature-changes of the surrounding air 424
Magnetic force at a point on its axis due to a current in a helical coil of one turn, The 900
Magnetic moment of the manganic ion, The 1029
Magnetic properties of amorphous manganese, The 197
Magnetic susceptibility of cerium chloride in aqueous solution and its variation with temperature 559
Magnetized surfaces, Demonstration of a simple apparatus for measuring the pull between two 186
Magnetron oscillator, Electronic theory and the 1
Mahanti, P. C.: The band spectrum of vanadium oxide 433
Manganese, The magnetic properties of amorphous 197
Manganic ion, The magnetic moment of the 1029
Manley, John L.: A new precision colorimeter 69
Marris, G. C., *see* Campbell, N. R.
Martyn, D. F.: The propagation of medium radio waves in the ionosphere 323

PAGE
Martyn, D. F., R. O. Cherry, and A. L. Green: Long-distance observations of radio waves of medium frequencies 340
McLachlan, N. W. and A. L. Meyers: Spherical sound-waves of finite amplitude 644
Mercury, Adsorption of gases on 460
Mercury arcs and their variation with temperature-changes of the surrounding air, Absolute intensities in the spectrum of quartz 424
Meyers, A. L., *see* McLachlan, N. W.
Microphotometric methods in divided-beam spectrophotometry, The use of . 125
Millington, G.: Ionization charts of the upper atmosphere 263
Minnis, C. M., *see* Sloane, R. H.
Morris, D. K., Obituary notice of 1136
Mott, N. F.: A discussion of the transition metals on the basis of quantum mechanics 571
Mutual inductance calculations, The computation of the integrals required in . 86

Negative glow in nitrogen, Note on additional experiments on the effective rotation temperature of the 413
Nethercot, W., *see* Whitehead, S.
Nettleton, H. R. and E. G. Balls: The absolute measurement of electrical resistance by a new rotating-coil method 54
Neutrons, The disintegration of boron by 873
Nickel in longitudinal magnetic fields, The longitudinal thermoelectric effect (2) 852
Nickel wire, Thermomagnetic hysteresis in 910
Nitrogen first positive hands by electron impact, The efficiency of excitation of the 420
Nitrogen, Note on additional experiments on the effective rotation temperature of the negative glow in 413

Oatley, C. W.: A negative-resistance device and its application to harmonic analysis 471
Obituary notices 1129
Observations of points connected by a linear relation, On 92
Oddie, T. H.: The efficiency of separation of hydrogen and deuterium by electrolysis 1060
Oscillograph by harmonics, Time-marking a cathode-ray 258
Oxide films on liquid metals studied by means of electron-diffraction . . . 109

Palmer, Professor L. S. and Denis Taylor: On a theory of the action of rectangular short-wave frame aerials 377
Palmer, Professor L. S. and Roy Witty: The current variations in a short-wave square frame aerial revolving in its own plane 388
Pantulu, D. V. Reddi, *see* Bates, L. F.
Particles in air, On the behaviour of suspended 136
Pearson, A. R. and B. Pleasance: The colour temperatures of the Hefner and acetylene flames 1032
Pearson, E. B.: On the behaviour of suspended particles in air and the velocity of sound at supersonic frequencies 136
Pearson, H., *see* Simon, F.
Pegler, Geoffrey D.: A dynamometer null method of measuring the inductance and the effective resistance of iron-cored chokes carrying direct current . 964
Peltier coefficient, The direct measurement of the 615
Penman, H. L.: The effect of temperature on supersonic dispersion in gases . 543
Petrie, D. P. R.: Comparison of X-ray wave-lengths by the plane-grating vacuum spectrograph, and the structure of the K line of carbon 626
Phosphorus monoxide, Rotational analysis of the ultra-violet bands of . . 247

PAGE
Photocell and colour filters, The colour-matching of tungsten-filament lamps by means of a single 1012
Photoelectric cell, A new selenium-sulphur rectifier 477
Photographic shutters, New method of measuring the time and efficiency of . 446
Pi, T. H. and William Band: The longitudinal thermoelectric effect: (2) Nickel in longitudinal magnetic fields 852
Pitt, F. H. G., *see* Wright, W. D.
Plates, An experimental determination of the frequencies of free circular . . 794
Pleasance, B., *see* Pearson, A. R.
Polarized light and Hertz's problem for a uniaxial material, Convergent . . 306
Potential acquired in the natural electric field by a vertical rod standing on the ground, insulated at the bottom and carrying a collector at the top, The . 218
Preston, J. S.: The colour-matching of tungsten-filament lamps by means of a single photocell and colour filters 1012
Pulley, O. O.; A receiver discriminating between right and left-hand circularly polarized wireless waves 1098

Quantum mechanics, A discussion of the transition metals on the basis of . . 571

Radio waves in the ionosphere, The propagation of medium 323
Radio waves of medium frequencies, Long distance observations of . . . 340
Radley, W. G., *see* Whitehead, S.
Railston, W. and E. G. Richardson: The effect of pressure on supersonic dispersion in gases 533, 777
Raman spectrum of a ferromagnetic oxide, A note on the 877
Rayleigh disc as applied to the measurement of sound-intensity in water, A correction to the theory of the 779
Resistance by a new rotating-coil method, The absolute measurement of electrical 54
Resistance of iron-cored chokes carrying direct current, A dynamometer null method of measuring the inductance and the effective 964
Resistivity of ground in the United Kingdom 923
Resistivity of the earth, Measurement of the 589
Reviews of books 191, 371, 521, 774, 1000, 1139
Richardson, E. G., *see* Railston, W.
Richardson, L. F.: Time-marking a cathode-ray oscillograph by harmonics . 258
Ripples as a diffraction grating, Demonstration of the use of a liquid surface carrying 998

Samuel, R., *see* Asundi, R. H.
Schuster, Sir Arthur, Obituary notice of 1130
Searle, G. F. C.: The magnetic force at a point on its axis due to a current in a helical coil of one turn 900
Seismology, Time and amplitude relations in 455
Selwyn, E. W. H., *see* Eyles, E. D.
Shutters, New method of measuring the time and efficiency of photographic . 446
Silver, Longitudinal thermoelectric effect 904
Simon, F., A. H. Cooke, and H. Pearson: Liquefaction of hydrogen by the expansion method 678
Sloane, R. H. and C. M. Minnis: Spectroscopic observations of recurrent phenomena in discharge tubes 1019
Smith, F. D., *see* Wood, A. B.

PAGE

Smith-Rose, R. L.: The electrical properties of soil at frequencies up to 100 megacycles per second; with a note on the resistivity of ground in the United Kingdom 923
Soil at frequencies up to 100 megacycles per second, The electrical properties of the 923
Sound at supersonic frequencies, The velocity of 136
Sound in sheet material, Demonstration of the velocity of 185
Sound in sheet materials, The velocity of 149
Sound-intensity in water, A correction to the theory of the Rayleigh disc as applied to the measurement of 779
Sound-waves of finite amplitude, Spherical 644
Spectrograph, Comparison of X-ray wave-lengths by the plane-grating vacuum . 626
Spectrographic examination of arcs between plain soot carbons and its connection with the candle-power per ampere of the positive crater, A 667
Spectrophotometry, The use of microphotometric methods in divided-beam . 125
Spectroscopic observations of recurrent phenomena in discharge tubes . . 1019
Spectrum of a ferromagnetic oxide, A note on the Raman 877
Spectrum of beryllium monoxide, The band 415
Spectrum of quartz mercury arcs and their variation with temperature-changes of the surrounding air, Absolute intensities in the 424
Spectrum of vanadium oxide, The band 433
Starr, A. T.: The rectifying peak voltmeter as a standard instrument . . . 184
Sulston, W. J.: The temperature variation of the viscosity of aqueous solutions of strong electrolytes 657
Supersonic dispersion in gases, The effect of pressure on 533
Supersonic dispersion in gases, The effect of temperature on 543
Supersonic frequencies, The velocity of sound at 136
Surface tension of a moving water sheet, The 549

Taylor, Denis, *see* Palmer, L. S.
Taylor, H. J.: The disintegration of boron by neutrons 873
Taylor, W. and H. W. Lee: The development of the photographic lens . . 502
Thermal expansion of cast and rolled zinc, Measurements of the . . . 1117
Thermoelectric effect, The longitudinal: (2) Nickel in longitudinal magnetic fields 852
Thermoelectric effect, The longitudinal: (3) Aluminium 859
Thermoelectric effect, The longitudinal: (4) A further study of aluminium . 862
Thermoelectric effect, Longitudinal: (5) Silver 904
Thermomagnetic hysteresis in nickel wire 910
Thompson, N.: Note on additional experiments on the effective rotation temperature of the negative glow in nitrogen 413
Tin, The twinning of single crystals of 733
Trichromats, The colour-vision characteristics of two 205
Triode, The electron-oscillation characteristics of an experimental plane-electrode 1042

Ultra-violet bands of phosphorus monoxide, Rotational analysis of the . . 247

Vanadium oxide, The band spectrum of 433
Viscosity of aqueous solutions of strong electrolytes, The temperature variation of the 657
Viscosity of liquids at high pressures, Demonstration of apparatus for measuring the 519
Voltmeter as a standard instrument, The rectifying peak 184
Vortex motion in gaseous jets and the origin of their sensitivity to sound, On . 703

PAGE

Whitehead, S. and W. G. Radley: Experiments relating to the distribution of alternating currents in the earth and the measurement of the resistivity of the earth 589
Whitehead, S. and W. Nethercot: The breakdown of dielectrics under high voltage, with particular reference to thermal instability 974
Williams, S. E.: The efficiency of excitation of the nitrogen first positive bands by electron impact 420
Wireless waves, A receiver discriminating between right and left-hand circularly polarized 1098
Witts, F. J., Obituary notice of 1129
Witty, Roy, *see* Palmer, L. S.
Wood, A. B.: A correction to the theory of the Rayleigh disc as applied to the measurement of sound-intensity in water 779
Wood, A. B.: An experimental determination of the frequencies of free circular plates 794
Wood, A. B. and F. D. Smith: The velocity of sound in sheet materials . . 149
Woodall, A. J.: The direct measurement of the Peltier coefficient 615
Wright, W. D. and F. H. G. Pitt: The colour-vision characteristics of two trichromats 205

X-ray crystal analysis, The absorption factor for the powder and rotating-crystal methods of 879
X-ray reflections from crystalline powders, The quantitative measurement of the intensity of 932
X-ray region, The theory of the formation of an image by a plane band grating used in the soft 948
X-ray wave-lengths by the plane-grating vacuum spectrograph, Comparison of . 626

Zaki-uddin, M., *see* Asundi, R. H.
Zinc, Measurements of the thermal expansion of cast and rolled 1117
Zinc oxide, The lattice-dimensions of 835

CAMBRIDGE: PRINTED BY WALTER LEWIS, M.A., AT THE UNIVERSITY PRESS

157